超循环论

当代学术思潮译丛



M·艾根
P·舒斯特尔 著



超循环论

著 者/〔联邦德国〕M·艾根
P·舒斯特尔
译 者/ 曾国屏 沈小峰

● 上海译文出版社

M. Eigen & P. Schuster

The Hypercycle

A Principle of Natural Self-Organization Springer-Verlag 1979

本书根据施普林格出版社 1979 年版译出

超 循 环 论

〔联邦德国〕 M・艾根 P・舒斯特尔著

曾国屏 沈小峰译

上海译文出版社出版、发行

上海延安中路 955 弄 14 号

全 国 新 华 书 店 经 销

上 海 译 文 印 刷 厂 印 刷

开本 850×1168 1/32 印张 14 插页 3 字数 321,000

1990 年 5 月第 1 版 1990 年 5 月第 1 次印刷

印数：0,001—4,000 册

ISBN 7-5327-0487-4/B・037

定价：7.15 元



本书作者M·艾恩

译者的话

（一）

20世纪以来，科学前沿的探索主要在无穷大和无穷小方面，研究宇宙之初，粒子之渺，以及高温、深冷等等各种极端条件下的自然现象。人们普遍认为，我们对自身及周围世界的规律似乎已经认识得比较清楚了，因此没有什么新的学说能使我们惊奇了，以致有人又一次宣布了“理论物理学的终结”！

但是，第二次世界大战以后，特别是50年代以来，人们发现，不仅在微观和宇观领域，而且就是在宏观领域，都仍然充满着许多意想不到的事情。对于生命的起源，生物的进化，人体的功能，思维的奥秘，乃至社会经济文化的变革这样一些复杂性事物的演化发展规律，人们实际上是知之甚少的，有许多新的奇妙现象无法用现有的科学去解释。目前，探索复杂性成了科学研究的另一个发展方向。

在探索复杂性的征途上，系统科学的出现和发展特别引人注目。40年代中相继建立了系统论、控制论和信息论，它们将自然、社会现象作为系统来研究其信息传递、控制反馈等方面的问题，这在理论和应用上都取得了很大成功。自此之后，又出现了许多以系统为研究对象的理论。

在当代众多的系统理论中，既有比较严密的数学、物理学理论基础，又有一定的实验依据，并在自然科学领域和社会经济文化生活中得到广泛应用的，当首推耗散结构理论、协同学和超循环理论。这三种关于非平衡系统的自组织理论在70年代前后相继建立不是偶然的，它们是当代科学在探索复杂性、建立系统科学的过程中的重要进展。

我国科学家钱学森在倡导建立系统科学的基础学科——系统学时，高度赞扬了普里戈金、艾根和哈肯等人的工作，他在几年以前就强调过，把包括他们自组织理论在内的工作收集和组织起来，是当务之急。最近，他在主持研究编写《系统学》时又说："我认为把运筹学、控制论和信息论同贝塔朗菲、普里戈金、哈肯、弗洛里希、艾根等人的工作融会贯通，加以整理，就可以写出《系统学》这本书。"因此，将艾根的超循环理论翻译出版，对于建立和发展系统科学有重要的科学意义。

（二）

艾根(Manfred Eigen)是一位音乐家的儿子，1929年5月9日出生在德国的波鸿。艾根青年时期

在格丁根大学学习物理学和化学。在他18岁生日那天，亦即德国法西斯正式签订投降书的第二天，他被征入高炮部队，从而一度中断了学习。战后不久，他重返格丁根大学继续学业，于1951年获得格丁根大学的博士学位。同年，他成为该大学物理化学研究所的助理员。1953年他进入了马克斯·普朗克物理化学研究所（即现在的马克斯·普朗克生物物理化学研究所），1958年成为该所的研究员，1962年进而成为该所的生物化学动力学室主任，从1964年开始担任该所所长。

艾根早年的研究工作主要在快速化学反应动力学及其反应机理方面。一个快速化学反应往往在百万分之一秒(10^{-6}秒)以内就完成了，因此以前在研究其动力学及反应机理上遇到了困难。1954年，艾根成功地把弛豫技术引入到快速化学反应研究中，为研究提供了一种新的方法，从而得到了一系列重要的结果。

正是由于快速化学反应研究的出色成就，导致了艾根与诺里什(R. G. Norrish，英)和波特(G. Porter，英)一起，荣获了1967年诺贝尔化学奖。不过，艾根主要是利用弛豫法研究了许多液相中的快速化学反应，而诺里什和波特则主要是利用闪光分解法

研究了许多气相中的快速化学反应。弛豫法现已成为化学反应动力学研究中的一种基本方法。

艾根在快速化学反应研究中，特别注意了生物体内发生的快速生物化学反应，并从生物分子演化的角度来对它进行考察。这使他对核酸和蛋白质的起源及其相互关系产生了兴趣，最终导致了他对生命起源的一个关键问题——生物信息起源的开创性探讨。

通过他自己的工作，同时也由于分子生物学、非平衡热力学、自组织现象研究、信息论和博奕论等方面进展的推动，艾根逐渐从实验和理论两方面步入了探索生命起源的领域。1970 年他在讲演中提出了超循环思想；1971年在德国《自然杂志》上发表《物质的自组织和生物大分子的进化》一文，正式建立了超循环理论。在1973年出版的世界著名科学家庆贺狄拉克70寿辰的《物理学家的自然概念》论文集中，他又发表了《生物信息的起源》一文，阐述和发展了他的思想。他与温克勒(R. Winkler)合作于1975年发表的《博奕论——偶然性的自然选择》一书，也以探讨生物大分子自组织为主题。他与理论化学家舒斯特（Peter Schuster）合作，于1977—1978年在《自然杂志》上连续发表了3篇系列论文，系统地阐

明了超循环理论；在此基础上，1979年他们整理出版了《超循环：一个自然的自组织原理》一书。本书即由《超循环：一个自然的自组织原理》、《物质的自组织和生物大分子的进化》和《生物信息的起源》三部分组成。这样可以使读者对于超循环理论有一个系统的了解。

艾根除了与人共享诺贝尔化学奖以外，他还单独获得了本生学会的波登坦奖、奥托·哈恩化学奖、美国化学会的林纳斯·鲍林奖章、瑞士化学会的帕拉塞斯奖章、红外科学研究基金会的红外科学奖等10来种科学奖。他是不少国家的科学学会、组织的成员或荣誉成员，这些组织中包括英国的法拉第学会、美国国家艺术和科学院、奥地利科学院、丹麦皇家科学院和苏联科学院等。他还获得了一些国际知名大学的荣誉职位和学位。

（三）

达尔文在19世纪中叶建立了生物进化论以后，一个多世纪过去了，生物进化论的研究已有了相当的进展，但是关于生命起源的问题却至今没有解决。恩格斯在19世纪70年代曾经预言："生命的起源

必然是通过化学的途径实现的。”(《反杜林论》，人民出版社，1970年版，第70页）苏联生物学家奥巴林于本世纪30年代从科学上倡导化学进化学说。50年代诞生了分子生物学以后，人们又在模拟原始地球的条件下，在实验室合成了构成生命的基础有机物——蛋白质和核酸。

但是，在理解核酸和蛋白质的关系上，又遇到了新的困难，这就是“先有核酸还是先有蛋白质”的问题，或者更抽象地说即“先有信息还是先有功能”的问题。这使人想起了古老的“蛋、鸡”之争。超循环理论探讨了这一问题。艾根指出，在生物信息起源上的这种“在先”，不是指时间顺序，而是指因果关系；事实上，提出“在先”的问题，不是提出了一个科学问题，而是一个伪问题。这里有一种双向的因果关系，或者说是一种互为因果的封闭圈。核酸和蛋白质的相互作用，相当于“封闭圈”即“循环”的一个复杂的等级组织。从反应循环，到催化循环，再到超循环就构成了一个从低级到高级的循环组织。

超循环理论的提出，也是对生物学中多样性和统一性关系深入思考的结果。地球上的生物有数百万种之多，其形态结构、生理机制和生态习性各异，因而存在着多样性。然而，在它们的细胞中又只有

一种基本的分子机构，即普适的遗传密码以及基本一致的翻译机构和一种大分子手性，其中翻译过程的实现又要求数百种分子的配合，因而又存在着统一性。艾根指出，这种统一性很难想象是一下子形成的，除非是一次巨大的创世行动的结果。他认为，在生命起源和发展中的化学进化阶段和生物学进化阶段之间，有一个分子自组织过程。因此，进化可以划分为如下几个阶段：

1．化学进化阶段；

2．分子自组织进化阶段；

3．生物学进化阶段。

在分子自组织进化阶段，既要产生、保持和积累信息，又要能选择、复制和进化，从而形成统一的细胞机构，因此这个自组织过程只有采取超循环的组织形式。

艾根认为，超循环组织和一般的自组织一样，它必定起源于随机过程，开始于随机事件，但是，只要条件具备，它的出现虽然不是决定论的，但却是不可避免的。在超循环自组织过程中，也包含了许多随机事件，但是，这些随机效应能反馈到它们的起点，使得它们本身变成某种放大作用的原因。经过因果的多重循环、自我复制和选择，功能不断

完善，信息不断积累，从而向高度有序的宏观组织进化。他认为，“上帝不掷骰子”和“绝对偶然”的观点在此都是不正确的。

艾根指出，达尔文自然选择原理不仅是生物学进化的原理，而且也是研究超循环自组织的指导性原理。达尔文原理强调“物竞天择，适者生存”，艾根在对超循环的研究中指出，在此不仅“生存竞争”、“空间隔离”是重要的，而且“协同作用”、“整合作用”同样是重要的。超循环组织，作为一个远离平衡的开放系统，既竞争又协同，既隔离又整合，从而选择和进化。因此，艾根的超循环理论在分子水平上把竞争和协同结合起来，发展了达尔文原理。

艾根强调，他之所以提出超循环理论，是想把物理学普遍原理推广到生物学并与生物学成果相结合，对经验事实进行抽象，从现存中追踪历史的遗迹，从而用模型反映现实的结果。正是在建立抽象模型的基础上，他运用了包括非线性微分方程、概率论、博奕论、不动点分析等数学工具对模型进行了定量的处理，从而得出定量的、规律性的富有意义的结果。艾根提出的超循环理论，虽然有一定的实验事实支持，但还有待于进一步改进和完善。他希望他的理论能够“激起……进行新的实验，从而有

助于更深刻地理解作为‘自然规则’的生命以及生命的起源。”

艾根还特别指出，在神经组织和社会组织中，也存在超循环的组织形式，超循环理论的许多结论，不仅具有自然科学意义，而且具有社会科学意义。

* * *

在翻译过程中，曾分别求教于漆安慎、李占柄、郭华庆、蔡兵等多位专家、学者，他们分别审阅了部分译文或提出了宝贵意见；翻译中还参考了徐京华先生发表的有关译文，以及李建华同志的有关译稿；卢治恒副教授协助翻译了书中一节德文；在西德任客座教授的安文铸博士协助约请艾根为本书写了序言并代译为中文；刘未、沈文、李莉、万伟协助抄写了部分稿件；谨在此表示衷心感谢。本书论述的是一门新的学科，涉及到数学、物理学、化学、生物化学、生物学、信息论等多学科知识和词汇，有许多新的科学名词术语，译者的专业和文字能力有限，虽然求教了许多有关专家，但翻译中错讹之处在所难免，敬请读者批评指正。

译者

1987.10

作者为中译本写的序

分子生物学的研究成果为达尔文的进化论提供了新的、物理学的牢固基础。核酸的复制动力学——一种借助于模型推演出任何生物繁殖所依赖的聚合过程——已经被当作某些重要的进化现象(如筛选、优化、对环境的适应或共生协同)进行定量描述的新的途径。分子进化论与传统的群体遗传学观点的主要区别在于以下两点：

1．变异是复制过程中不可缺少的组成部分。不论是无缺损的还是有缺损的复制都是一个或同种反应机制的平行反应。借助于这种新的观点，人们有可能对于频繁变异系统进行没有局限的分析。

2．决定筛选过程结果的那些量(如，在群体遗传学中的适合度)获得了建立在物理-化学特性基础上的一个新的解释，这些物理-化学特性（至少在原则上）可以不依赖于筛选过程而被测量出来。达尔文理论中的同语反复问题(即经常借助“适者生存”表达的循环推理)从而简单地被解决了。在特别简单的系统(例如，用在试管中的RNA分子所作的进化实验或存在于体内和体外的简单病毒)中，一种对进化现象的定量的实验途径已获成功。

分子进化理论发展了一些新的概念，这里给出两个例子：

“拟种”被理解为一个群体中稳定的变异分布；“超循环”是在自复制元素中的一个有组织的全体，在这个整体中，导致筛选的竞争通过相互依赖的简单形式被联结在一起。呈现在读者面前的这个中文文本包含着分子进化理论在1979年时的概貌，其中《超循环——一种自然的自组织原理》汇集了1977年和1978年在《自然科学》杂志上发表的三篇论文。尽管从那以后，该理论继续有了发展，但是其基本理论在《超循环》发表以后只有极少的变动。因此，从今天看来，这本书仍可视为超循环理论从一个简单的观点到新的理论的入门。不考虑理论的新发展，这本入门著作首先是包含了大量的实验结果，同时介绍了新的进化理论的许多应用可能性。借此，我们在这里附上一个新成果的出版物索引。

Literaturhinweise:

[1] M. Eigen, J. McCaskill, P.Schuster. *Adv. Chem. Phys.* 75 (1988) ****.

[2] M. Eigen, P. Schuster. *J. Mol. Evol.* 19 (1982) 47-61.

[3] M. Eigen. *Stufen zum Leben.* Piper Verlag, München 1987.

[4] M. Eigen, W. C. Gardiner, P. Schuster, R. Winkler-Oswatitsch. *Sci. Am.* 244(4) (1981) 88-118.

[5] M. Eigen. *Chemica Scripta* 26B (1986) 13-26.

[6] M. Eigen. *Ber. Bunsenges. Phys. Chem.* 89 (1985) 658-667.

[7] P. Schuster. *Physica Scripta* 35 (1987) 402-416.

[8] P. Schuster, K. Sigmund. *Ber. Bunsenges. Phys. Chem.* 89 (1985) 668-682.

[9] C. K. Biebricher. *Evolutionary Biology* 16 (1983) 1-52.

[10] C. K. Biebricher, M. Eigen, W. C. Gardiner. *Biochemistry* 22 (1983) 2544-2559.

[11] C. K. Biebricher, M. Eigen, W. C. Gardiner. *Biochemistry* 23 (1984) 3186-3194.

[12] C. K. Biebricher, M. Eigen, W. C. Gardiner. *Biochemistry* 24 (1985) 6550-6560.

[13] C. K. Biebricher, M. Eigen, R. Luce. *J. Mol. Biol.* 148 (1981) 369-390.
[14] C. K. Biebricher, M. Eigen, R. Luce. *J. Mol. Biol.* 148 (1981) 391-410.
[15] C. K. Biebricher, S. Diekmann, R. Luce. *J. Mol. Biol.* 154 (1982) 629-648.

Mit besten Grüssen:

M. Eigen

Göttingen, 1. Juli 1988.

1988年7月1日于格丁根

(安文铸 译 卢治恒 校)

目次

超循环：一个自然的自组织原理

艾　根
舒斯特

前言

本书取材于发表在1977—1978年《自然杂志》上的一系列论文。从逻辑结构上考虑，全书分成三个部分，即可概括为以下三个论点：

1．超循环是一个自然的自组织原理，它使一组功能上耦合的自复制体整合起来并一起进化。

2．超循环是一类全新的、具有独特性质的非线性反应网络，它便于进行统一的数学处理。

3．超循环可以通过趋异突变基因的稳定化，而起源于某种达尔文拟种的突变体分布中，一旦聚集起来，超循环将经一个类似于基因复制及特化的过程，进化到更复杂的程度。

为了扼要说明第一个论点的意义，我们可以参考另一个物质自组织原理——达尔文自然选择原理。正如我们现在所看到的，这一原理描述了有关产生信息的唯一为人们所理解的手段，不管这种信息是复杂的活有机体的蓝图或是一种有意义的字符序列，这种复杂的活有机体是从较不复杂的祖先形式进化而来的，而这种字符序列的选择可以用进化模型的博奕模拟。

自然选择（这里强调的是“自然”一词）以自复制为基础。这就是说，对于由有限的普通物质组成的自复制体系统，自然选

择将是不可避免的结果。同样地，由自然选择支配的进化行为，其基础是有干扰的自复制。这些物理性质足以保证可重复地形成高度复杂的系统，即产生可作为活的有机体蓝本的信息。然而，所要获得的信息量有一定的限度，这是达尔文自然选择机制所固有的。这正是超循环登台的地方。超循环也是一个自组织的原理，只是前提不同，因而所产生的结果也不同。

第一部分概括的达尔文系统，主要表明了两个结果：

(1) 自复制体为选择而竞争。对于落入小生境的无联系的物种，这种竞争可以得到缓和。然而，为保持野生型的稳定，它必须在每一种突变体分布中都起作用。没有这种竞争稳定化，其信息便会丧失殆尽。

(2) 稳定的野生型信息量是有限的。换言之，信息量必须保持在某个阈值之下，阈值的大小与(每符号的)平均误差率成反比。阈值还依赖于野生型的优势的对数，此优势即相对于全体(稳定)分布的突变体的平均选择优势。一旦出现了某种突变体，该分布就会变得不稳定，因为它扰乱了这种有利于先前稳定的野生型条件。

这些性质是达尔文系统所固有的。它们保证了进化行为，进化行为的特征是选择和最适自复制体的稳定复制，并且它可被适应性更强的任何突变体取而代之。另一方面，这种系统的进化被限制在由最大信息量的阈值所确定的某一复杂性水平上。

由于这种限定，最初的自复制体必须是相对短的核酸链。只有这类大分子才充分满足内在的自复制条件。然而，作为自复制精确性基础的物理力，其专一性是有限的。只有催化支持才能增进精确度，在此，为了适应进化，催化剂也必须是可复制的。在这个进化阶段需要信息翻译，该信息是由复制物质经遗传得到的。

要跨越这一步是极其困难的。进化必定几乎停止了。这里需要某种机构，而为了产生它，又必须立刻利用这一种机构。哪怕是原始的翻译器，也至少要包括分配4种不同的氨基酸以及相应数目的酶和信使的4种匹配器。这种系统所需的信息量，相当于单链*RNA*病毒的信息量。不过，这些微粒能够利用其宿主细胞的完善的翻译器。它们在高度适应的酶机构的帮助下进一步复制了由酶机构所代表的某种最终的、也是优化的进化产物。

*RNA*噬菌体的基因组很少超过几千个核苷酸，这些核苷酸足以对几种(例如4种)蛋白质分子进行编码。正如第一部分表明的，这种限制是由只在高度适应的复制酶帮助下才能达到的精确度决定的。信息量的任何进一步扩充，都将需要如此精致的机理，如校正读码，其中包括核酸外切酶和连接酶的作用，而只有处于相当高级进化阶段的*DNA*聚合酶才能利用这种机理。如果复制精确度仅是以核酸所继承的物理性质为基础，而不允许在任何一条核苷酸链上复制积累起多于50至上百个核苷酸，那么即使是原始的翻译系统，又怎么能够产生呢？一个翻译系统所必需的信息量相当于自复制单链中可得到的信息量的几倍。

超循环是一种工具，它把那些长度有限的自复制体整合到某种新的稳定序中，从而能够相干地进化。任何其他种类组织，诸如仅有空间隔离的组织或者非循环网络，都不可能同时满足如下3个条件：

1) 为保存它们的信息，要在每一自复制体的野生型分布中保持竞争；

2) 允许几种(除了竞争的)实体及其突变体分布共存；以及

3) 把这些实体统一成某个相干的进化单元，其中每一个体的优势都能够被所有成员加以利用，而且这个单元在此作为一

个整体，在与任何可选择的组分单元的激烈竞争中都得以继续存在。

我们的陈述包含了第一个结果，它代表了一种逻辑推论：

如果我们寻求某种保证翻译器连续进化的物理机制，那么超循环组织就是一个最起码的要求。其中的信息载体具有自复制性质在此是必要的，但并非充分的。若我们分析超循环组织的条件，则立刻可以发现它们相当于达尔文选择的前提。达尔文选择的基础是自复制，它是一种线性自催化。正如第一部分表明的，超循环是自催化系统等级结构中更高的层次。它是由自催化或复制循环所构成的那些复制循环是被循环的催化联系起来的，即被另外的重叠的自催化联系起来的。所以，超循环是以(二级或更高级的)非线性自催化为基础的。

超循环可以当作一类特殊的反应网络加以分析，因为它们表现出"规则的"行为。所以在第二部分中进行了这种一般的分析(参见陈述2)。它们表现出独特的、其他类型耦合不具备的性质，从而可作为"抽象的超循环"加以统一处理。对这题材作如此解释表明它是一种很好的描述。

另一方面，超循环决不仅仅是我们心灵的抽象产物。该原理在*RNA*噬菌体的感染过程中仍然被保留着，尽管在此它被运用于宿主细胞的封闭世界中。此噬菌体基因组对翻译提供了一个因子，用作复制酶复合体的一个亚单元；复合体的其余部分由宿主因子来补充。这个噬菌体编码因子使得该酶具有绝对的噬菌体专一性。不考虑来自宿主源的全部*RNA*，噬菌体专一的复制酶复合物现在代表了噬菌体基因组的自催化放大的一个重叠反馈环路。

我们关于原始翻译器的超循环组织的必要性陈述，具有"如果—那么"的本性，而未涉及历史上实际存在的事物。在那里，

也许会出现意外的单个事件，即并不代表自然界的任何规则性的那些涨落，从而影响历史途径。如果我们想要表明，历史的进化确实是在某个特殊的物理原理支配下进行的，那么我们就必须找到历史证据，即存在于现在的有机体中的早期组织形式的遗迹。第三部分讨论了这个问题，我们的第三个陈述也谈到这个问题。

转移*RNA*作为翻译中的关键物质，提供了一些关于它们起源的信息。它们似乎提供了一种自然的方式，解决了非线性网络即集结问题的困难。网络中的所有成员都是同一主导拷贝即某个*tRNA*前体的后代。这种前体的拟种分布的突变体，可以在某个超循环组织原理起作用之前积累起来。由于与突变体密切相关，所以连接物和信使以及它们的翻译产物提供了非常相似的功能(作为目标和作为执行因子)，从而自动地"落入"包括一个循环的某个高度交联的组织中。正如第三部分表明的，通过耦合的进化特异性，这个循环就能够逐步地自我稳定，这种耦合可能是复制酶—靶型的，它们仍然被*RNS*—噬菌体利用着。此种现实的超循环由实验来检验，其中包括详细地研究现存的翻译机制。

我们希望，对于进化难题的研究，本书有助于提出正确的问题。任何理论，若其推论不能由实验核实，则其中就不存在绝对的价值。另一方面，理论不能只满足于解释实验事实，还必须提供更多的东西。正如爱因斯坦所说的：只有理论才能告诉我们哪些实验是有意义的。

在这个意义上，本书不仅是为探求把物理学规律统一地应用于自然界的物理学家们写的，它也是为化学家、生物化学家和生物学家们写的，以此激起他们进行新的实验，或许能为作为"自然规律性"的生命及其起源提供一种更深刻的理解。

与弗朗西斯·克里克、斯塔利·米勒和莱斯利·奥格尔的讨论，大大推动了我们的工作；这些讨论对我们来说意味着某种“选择压力”，以追求分子进化中的更多的连续性。克里斯托弗·比布里切尔、欧文·爱泼斯坦、伯恩特·格特、迪特马·波尔施克、卡尔·西格蒙德、保罗·伍利和罗伯特·伍尔夫的建议和批评，尤其使我们受益匪浅。鲁思尔德·温克勒-奥斯瓦提设计了本书中大多数插图，并总是讨论会的一位耐心的、关键性的参加者。

谨向一切为本书提供过帮助的人们致谢。

曼弗雷德·艾根
彼得·舒斯特
1978年9月6日
于格廷根

1 超循环的出现

1.1 进化中的统一性和多样性规范

动植物的种类有数百万之多，而其细胞又只有一种基本的分子机构：一种普适的遗传密码以及大分子的唯一手性，这是为什么？

现代的遗传学家会毫不犹豫地回答这个问题的前半部分：进化过程中有无数的复制和突变步骤，因而物种的多样性是进化过程的惊人的分支结果。这包括进食共同食物源的竞争者之间的选择，但也可以是孤立的或落入小生境的，也可以甚至是在足够温和的选择约束条件下的互容和共生。在优化适应环境方面，达尔文自然选择原理表述了一个指导性原理，以鉴别评估基因群体。在某种严格意义上，它只有在合适的边界条件下才是有效的，而这些条件在自然界可能得到满足，也可能得不到满足。在伟大的群体遗传学派的费希尔、霍尔丹和赖特的著作中，以精确的公式表达了自然选择原理，揭示了它的能力和局限。这样，该原理以活有机体这个先决条件为基础，尤其是建基于它们的复制机制。其中包括一些因子，它们解释了遗传的均一性和差异性，这是在尚不详知遗传的分子机制时业已认识

到的(表1.1)。

表1.1　自然选择的因子

遗传均一性因子	遗传差异性因子
基因复制	基因突变
基因聚合	聚合物的随机断裂
有丝分裂	染色体畸变
接合	减数(减数分裂)
连锁	交换
群体大小的限制	杂交
环境压力	个体适应性
子群体间的杂交	群体的再分
个体适应能力	亚群的局部环境

认识到生物世界的这种差异性的同时，事实上还存在一个怎样理解它在亚细胞水平上的均一性问题。许多生物学家只是把所有的前细胞进化事件合在一起,并将其看作“生命的起源”。的确，如果这曾是一次巨大的创世行动，它作为超出所有物理学统计期望值的独特的唯一事件而只发生过一次，那么我们确实能弄明白这种解释了。试图进一步理解这个“怎样”,则是枉费心机。偶然性只能归结为偶然性。

但是,我们关于哪怕是最简单细胞的分子精细结构的知识,也全然不支持这种解释。在这种非常复杂结构的构成中，其规律性明确地表明:第一个活细胞必定是长期进化过程的产物。在进化中必须包括许多单个的、但不一定是独一无二的步骤。特别是遗传密码，好象是这种多步进化过程的产物，很可能它从只由几种丰度最大的原始氨基酸独特的分配开始。尽管密码并未显示出那些最终分配方面的全部逻辑结构，但是它根本不是

随机的，人们不可能不得出那里有某种最优化原理在起作用的印象。人们可以把它称作最小变化原理，因为密码的结构是这样的，单个点突变的结果被归结为在氨基酸水平上的最小变化。冗余密码子，即对相同氨基酸编码的那些三联体，出现在邻近的位置上，而具有类似相互作用的氨基酸通常三个中只有一个不同，且往往出现在密码子的始端或末端。这种最优化作用，为了在进化过程中变得有效，需要通过反复试验，检验许多可选择对象，其中包括相当多的简并分配。因此，假若前细胞进化受到类似于达尔文自然选择机制支配的话，那么它的特征必然与我们在种群水平上所发现的分支程度相类似。

不过，对于遗传密码，哪怕在其精细的结构中，我们也没有遇到任何可选择的对象。下面的假设是很不符合要求的，如果假定最优分配总是碰巧只出现一次，并且发生在恰当的时候，此时又不容许任何选择，无疑这类选择会导致密码分支并引入不同的精细结构。另一方面，祈求决定论的物理事件唯一确定前细胞进化的历史途径，那也是难以令人信服的。

我们的研究结果表明，在达尔文物种进化的前面，还有一个类似的分子进化的渐进过程，由此导致了唯一的一种运用普适密码的细胞机构。这种密码最终确立起来，并不是因为它是唯一的选择，而是由于一种特殊的"一旦—永存"选择机制，这可以从任何随机分配开始。一旦建立则永存下去的选择，是超循环组织的一种结果。对大分子复制机制的详细分析表明，对于能够积累、保存和处理遗传信息的大分子组织来说，催化超循环是一种起码的要求。

1.2 什么是超循环？

考虑一个反应序列，其中的每一步产物，无论是否有附加的反应物的帮助，都要经历进一步的转变。如果在这个反应序列中，所生成的产物与上一步的一个反应物相同，那么此系统类似一种反应循环，这个循环在整体上就相当于一个催化剂。在最简单情况下，某一个分子例如酶代表了此催化剂，它使得底物转变成为产物：

$$S \xrightarrow{E} P$$

该形式系统的机制至少要求是3元循环(图1-1)。图1-2和

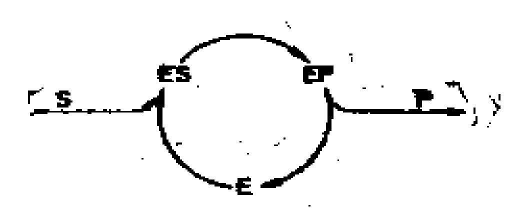


图 1-1　**酶的共同催化机制**　按米凯利斯和门顿的看法，其中(至少)含有3个中间物：游离酶(E)，酶-底物(ES)和酶-产物复合体(EP)。此模式表明，在从底物(S)到产物(P)的转变中，酶的催化作用等价于中间物的循环复原。当然，它只是对真实机制的一种形式表示，真实机制中可能涉及到底物的逐渐活化以及酶的诱导构象变化。

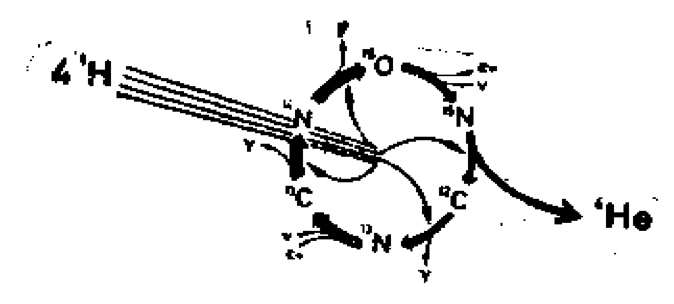


图 1-2　**碳循环**　这是贝特和韦察克尔提出的，它至少部分地相应于巨大恒星中的[illegible]。循环反应中[illegible]稳地重新产生出如下成分：^{12}C、^{13}N、^{13}C、^{14}N、^{15}O和^{15}N。[illegible]作为一个整体代表一种催化剂，它使得4个^1H原子转变成1个^4He原子，同时以γ量子、正电子e^+和中微子ν的形式释放能量。

图1-3所示的两个更复杂的反应循环，也满足基本的催化功能。贝特-韦泽克尔循环（图3-2）本质上适用于在巨大恒星中的高速率能量产生。可以这样说，它保持了太阳的光明，从而成为地球上生命的一个重要的外部前提。图1-3所示的克雷布斯循环即柠檬酸循环同样非常重要，它涉及到生命的内部机制。

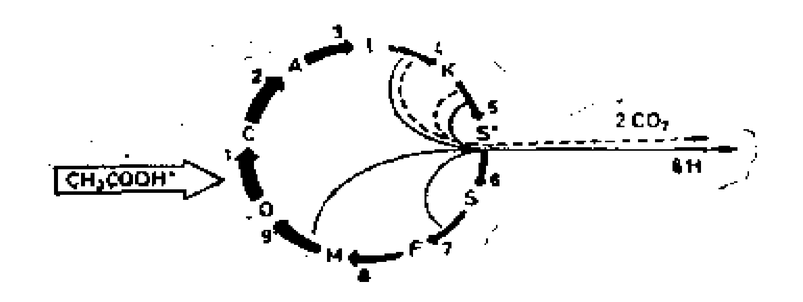


图 1-3　三羧酸循环（柠檬酸循环）　是燃料分子生物氧化的共同催化工具。其完整的系统是由克雷布斯提出的，圣捷尔吉、马修斯和努普对此也有重要贡献. 此循环中包括：柠檬酸(C)，顺乌头酸(A)，异柠檬酸(I)，α-酮戊二酸(K)，琥珀酰辅酶A(S*)，琥珀酸(S)，1-苹果酸盐(M)和草酰乙酸盐(O)。醋酸盐以激活的形式如乙酰辅酶A(步骤1)进入循环，并与草酰乙酸盐和水发生反应，形成柠檬酸(C)和辅酶A($+H^+$)。整个转化中，包括酶以及辅因子如辅酶A(步骤1,5,6)、F^{2+}(步骤2,3)、NAD^+(步骤4,5,9)、TPP、硫辛酸(步骤5)和FAD(步骤7)。附加的反应物H_2O(步骤1,3,8)、P和GDP(步骤6)，以及反应产物H_2O(步骤2)、H^+(步骤1,9)和GTP(步骤6)，在图中没有明显标出。净反应由两个乙酰碳完全氧化成CO_2(以及H_2O)构成。它形成了12个高能磷酸键，其中1个是在循环中形成的(GTP，步骤6)，其余11个是在NADH和$FADH_2$[3对电子对转移到NAD^+(步骤4,5,9)，1对转移到FAD(步骤7)]的氧化中形成的。

注：此循环作为一个整体起着催化剂作用，原因在于底物中间物的循环复原，但它不类似图1-4中描述的催化循环。尽管这个循环中每一步骤都是酶催化的，但是这些酶都不是通过循环形成的。

CoA=辅酶A，NAD=烟酰胺腺嘌呤二核苷酸，GTP=鸟苷三磷酸，FAD=黄素腺嘌呤二核苷酸，TPP=焦磷酸硫胺素，GDP=鸟苷二磷酸，P=磷酸

这个循环反应调节活细胞中的碳水化合物和脂肪酸的代谢过程，同时在合成代谢(即生物合成)过程中也有十分重要的作用。在两个模式中，根据守恒法则，高能物质转变为低能物质，亦即基本的中间物循环复原。这两个循环，尽管在起因上几乎毫不相干，在历史上却是几乎同时提出来的(1937—1938年)。

中间物单向地循环复原，表明了系统是远离平衡态的，并总是伴随着能量消耗，其中一部分能量耗散到环境中。另一方面，在封闭系统中出现的平衡，将使得每一步骤都是逐一平衡的。此种封闭系统中的催化作用将是微观可逆的，即在两个流向之上是同样有效的。

现在，让我们重复上述程序，以建立反应循环的等级结构并说明其特性。这意味着，我们在下一步考虑一个反应循环，其中，至少有一个中间物，但也可能是所有的中间物，其自身是催化剂。注意这些作为催化剂的中间物，它们在反应期间一个个都保持不变。它们一个个都是由高能原料流在前面的中间物的催化作用帮助下形成的(图1-4)。这种包含了大量中间物的系统，其成分必将是相当复杂的，因此在自然界中是少见的。最

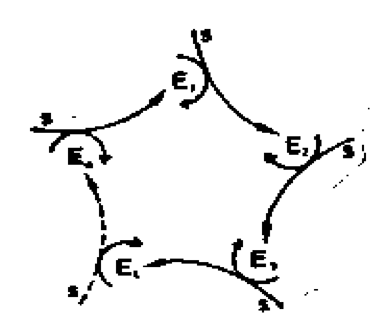


图 1-4　催化循环　在催化模式的等级组织中代表了一种较高水平的组织。循环中的组分 $E_1 \to E_n$ 自身是催化剂，它们从某些高能底物S生成，其中每一中间物 E_i 均是生成 E_{i+1} 的催化剂。催化循环被看作一个实体，等价于一个自催化剂，它指令自身的复制。要成为一个催化循环，只要所生成的中间物中有一个是后继的某一步反应的催化剂就可以了。

为大家熟悉的一个例子，是与RNA分子模板定向复制有关的4元循环(图1-5)。已在试管内对这种机制进行了研究，研究中运用了合适的反应介质并且用4种三磷酸核苷来缓冲，并运用一种噬菌体复制酶作为恒定的环境因子(库珀斯将给出更为详细的描述)。两链中的每一链都可以作为模板，指令其互补拷贝的合成，这类似于照相复印过程。

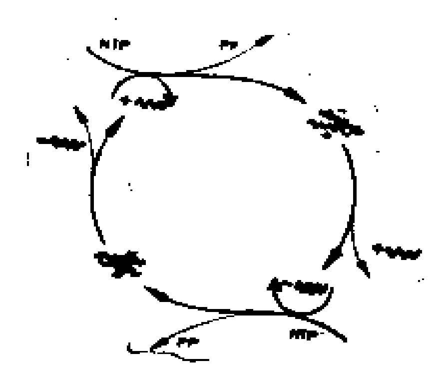


图 1-5　一个有生物重要性的催化循环　它由单链 RNA 的复制机制提供。其中的正链和负链作为相互复制的模板中间物。三磷酸核苷 (NTP) 作为高能原料，焦磷酸 (PP) 是代谢产物。互补指令代...了固有的自催化功能即自复制功能，我们将联系图1-11讨论其机制。

一种催化剂，或者某一整类信息载体I_i即自复制单元，是这类反应系统的最简单代表。此种过程可以形式地写为：

$$X \xrightarrow{I} I$$

本文将常考虑这类反应，我们用符号①来标记它们。

与单链RNA比较，双链DNA是这种真实的自复制形式，其中两链都同时通过聚合酶复制(参见图 1-6)。此种形式的模式适合于原核细胞，在原核细胞中遗传基本上限制于单个细胞系。

对于缓冲的底物浓度，简单催化系统和自催化系统共享一个速率项，它是催化剂浓度的一次项。不过，生长曲线将清楚地将两种系统区别开。条件固定时，简单催化过程的产物将随

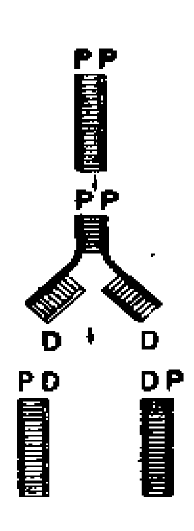


图 1-6　一个真实的自复制过程　以DNA复制中的一元循环为例，复制机制是相当复杂的(参见图1-12)，它保证了每条子链(D)与其中一条母链的结合。

时间线性地生长，而自催化系统却指数地牛长。

严格说来，自催化系统已可以称作超循环系统，因为它呈现出催化剂的循环排列，而这些催化剂自身又是反应循环。然而，我们将把这个术语限定于关于催化功能的超循环系综。它们实际上是二级或二级以上的超循环，因为它们至少是二级反应的催化剂浓度。

一个催化的超循环，是一个其中的自催化或自复制单元通过循环联接而联系起来的系统。图 1-7 示意了这种系统。作为自复制单元的中间物I_1到I_n，它们自身是催化循环，例如图1-5所示的RNA分子的正链和负链的结合。不过通过不同复制单元之间的特定耦合必须直接或间接地推动这种复制过程。更为现实的情况是，前一个RNA循环的翻译产物蛋白质可以影响这种耦合(图 1-8)。这些蛋白质可以作为特定的复制酶或去阻遏物，或者作为抵制降解的特定的保护因子。在自复制循环之间的耦合必定形成一种重叠的循环，于是只有整个系统才象一个超循环。

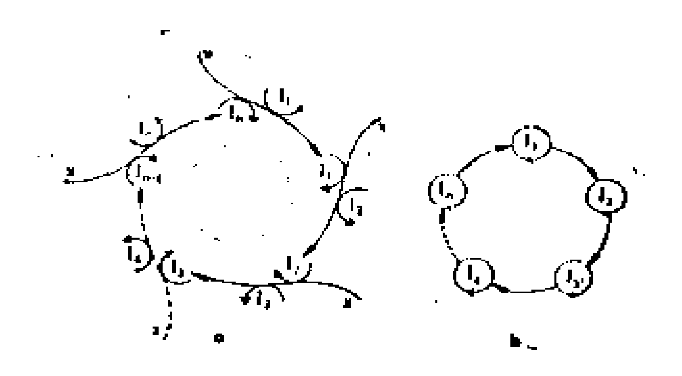


图 1-7　催化的超循环　由具有双重催化功能的自我指令单元I_i构成。作为自催化剂，或更一般地说作为催化循环，中间物I_i能够对它们自身的复制提出指令，另外又为下一个中间物（运用高能原料x）的复制提供了催化支持。简化了的图b示意了循环的分级结构。

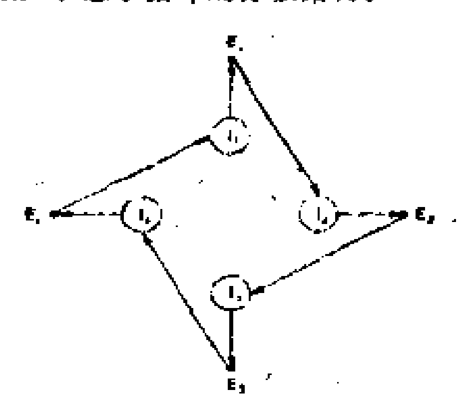


图 1-8　现实的二级超循环模型　其中信息载体I_i显示出两种指令，一种指令用于它们自己的复制，另一种指令把它们翻译成带有最优功能性质的另一类中间物E_i。每一个酶E_i都对于下一个信息载体I_{i+1}的复制提供催化帮助。它还可以包含进一步的，与翻译过程、代谢作用等有关的催化能力。在这样的情况下，超循环耦合是二级以上的。

把图1-4和图1-5进行比较，超循环是强化了的自复制。

这种类型的最简单代表，又是(准)一步系统，即强化了的自催化剂。我们在RNA噬菌体感染中遇到了这种系统(图1-9)。如果将噬菌体RNA(正链)注入某个细菌细胞中，那么它的基因型信息则被宿主细胞的机构加以翻译。于是，其中的一种翻译产物与一定的宿主因子结合起来形成某种活性酶复合物，此种

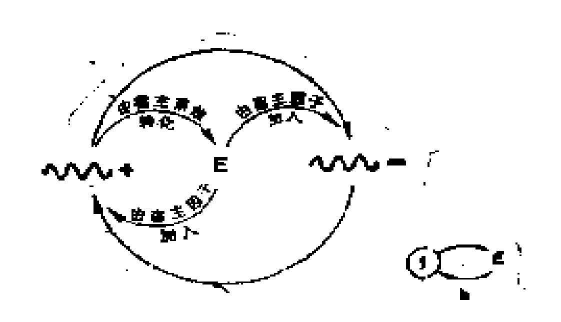


图 1-9 **RNA 噬菌体感染细菌细胞** 其中包含一个简单的超循环过程。运用宿主细胞的翻译机构，感染的正链首先指令合成一个蛋白质的亚单元(E)，E 与其他宿主蛋白质结合从而形成一种噬菌体专一的 RNA 复制酶。这种复制酶复合物专门识别噬菌体 RNA 的表现型特征，这是由正链和负链所展示出来的，因为在 RNA 链的特定区域具有对称性。其结果是噬菌体RNA的产生突然增加，由于超循环的本性，RNA的产生遵循双曲线生长定律(参见图1-17)，直到其中一种中间物达到饱和或者宿主细胞的代谢原料被耗尽为止。图b表示，如果其中一种中间物具有自催化或自我指令功能，则只要假定其余中间物通过封闭的循环联接反馈到它上面便足够了。

复合物能专门复制作为它们相互复制的模板的噬菌体RNA的正链和负链。不过，此复制酶复合物并不在任何可观的程度上显示出能增殖宿主细胞的信使RNA。感染的结果是出现了噬菌体颗粒的双曲线生长，但由于宿主细胞的资源终归是有限的，所以这种生长最终也是有限的。

对于孟德尔群体，只要群体数较少，在物种形成的起始阶段可以出现另一种自然的超循环。基因的复制要求两个等位基因(M 和 F)之间有相互作用，即在雄性和雌性染色体的同源区有相互作用，然后才会出现子代中的重排组合。孟德尔群体遗传学中，在速率方程(这导致双曲线生长而不是指数生长)里通常反映不出超循环的非线性性质，原因在于相对小的群体中出现了饱和，此时出生率(通常)变得只与群体中雌性数成比例。

正如图 1-10 的对照模式图中所见，超循环代表了组织的一个新层次。它们的独特性质表明了这个事实。非耦合的自复制单元保证了数量有限的、能够从一代传给下一代的信息保守性。业已证明，这是达尔文行为即选择和进化的一个必要前提。类似地，催化的超循环也是选择性的，但是，它们还具有整合的性质，允许与其他竞争单元协同作用。不过，它们与任何其他不是它们自身一部分的复制体的竞争，甚至比达尔文物种还要激烈。它们还有能力建立起作为它们一旦-永存型选择行为之结果的综合形式的组织，它不允许与其他超循环系统并存，除非它们是由更高级的联接而稳定化的。

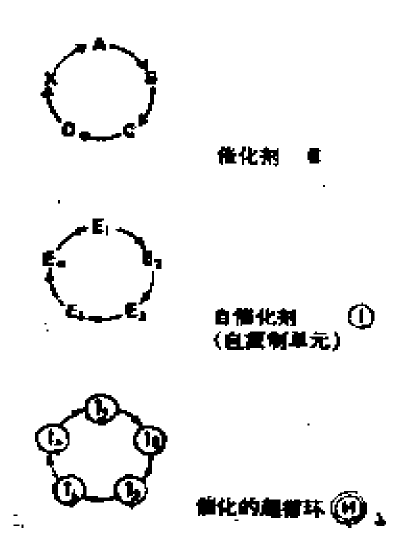


图 1-10　循环反应网络的等级结构　清楚地显示在这些相互对照的示意图中。图中：→表示化学转变，→表示催化作用。

超循环中最简单形式的耦合由直接的启动或去阻遏来表示，它们在速率方程中引入二级生成项。还可以出现更高级的耦合项，以确定超循环组织的组织度P。

单个超循环也可以联接起来，构成等级结构。不过，这需要严格依赖于组织度的内循环耦合项。为了建立稳定的共存，若

超循环H_1和H_2具有内组织度P_1和P_2，则要求其组织度为P_1+P_2的内循环耦合项。

本文的目的就在于：详细地阐述一种理论，这种理论对反应网络进行研究，我们将它命名为超循环。另一个目的在于讨论超循环在生物的自组织中的重要意义，尤其是在翻译的起源方面，它可能被看作是前细胞进化中的最决定性的步骤。

1.3 达尔文系统

1.3.1 自然选择原理

我们知道，在物理学中，原理不能还原为任何更基本的定律。作为公理，它们是来自经验的抽象，它们的预见应与可被实验检验的结果一致。热力学第一定律和第二定律就是典型的例子。

达尔文自然选择原理不能划入第一原理的范畴。正如群体遗传学中表明的，在一定外部条件约束下，自然选择显然是活的有机体基本性质的结果。所以，关于适者一词的意义，此原理把适者与其环境联系而作出了精确的断言，决不只是提出同义反复即所谓的"生存者生存"。即使把这个原理应用于具有可变的、且往往是未知边界条件的自然群体，这个原理仍然可提供进化事实的线索以及物种间系统关系的线索。这就是达尔文及其同代人华莱士的主要目标，即他们要为后人的信念提供一个更令人满意的基础。

实际上，现代群体遗传学的大多数工作，都是关心诸如孟德尔群体中遗传信息的扩散等更为实际的问题，而把诸如"活着

的"是否真的是选择和进化行为的必要前提这种学术问题置于一旁不顾。诸如代谢、自复制、有限的寿命以及可突变性等活有机体的明显属性，已可以解释在一定约束下的选择和进化行为，这就使得许多现代遗传学家相信，那些性质为生命现象所特有，它不可能在无生命世界中找到。在试管中进行的十分类似于自然选择和进化效果的实验，被解释为后生物学的发现，而不是解释为显示了典型的、特殊的物质行为。在此人们应该看到，甚至激光模也显示了自然选择现象；对它们的放大机制进行分析，揭示了比形式类比更多的东西。然而，不会有人用任何标准的定义把激光模说成"活的"。

只关心实际上活着的有机物的性质时，这些问题无关重要，但是，当这些问题与生命起源问题联系在一起，它们就有了极其重要的意义。为了找到适合于某个进化的自组织的那些分子系统，我们的确必须探寻必要前提。我们在大分子组织水平上遇到的潜在的复杂性，要求这个过程受到类似于选择和进化原理的支配，这些原理适用于生物界。

(上述引文中的)最近在理论和实验两方面的工作，已经涉及到这些问题。下面，我们将扼要说明前人关于达尔文系统的一些结果。

1.3.2 达尔文系统的必要前提

选择和进化的分子基础是什么？显然，这种行为并不是任意形式物质的综合属性，而是必须加以说明的物质的特殊属性的结果。一个系统若是自我选择的，则最基本的要求是，它必须以牺牲其他结构为代价使一定的结构稳定化。这种稳定化的标准本质上是动力学的，因为选择哪一个物种是由竞争者在任

一时候的分布决定的。换言之，任何结构，一旦被选择，就绝不存在静态的稳定性，只要出现了其他更“合适”的结构，或者处于新的环境中，它就可以变得不稳定。评价的标准中必须包括某种反馈性质，它保证了价值和动力学稳定性的一致性。某种有利的突变体，一旦作为某种涨落的结果产生出来，它就必定能够在存在着大量劣势竞争者的情况下放大自己。所以，优势必定至少与一些动力学性质是一致的，那些动力学性质负责放大作用。只有这样，系统才能在不存在“外部的选择者”时，有选择地组织自己。所需要的反馈性质，由内在的自催化即自复制行为所体现。

在运用博奕模型的一般分析中，对于在分子水平上产生出达尔文行为所必需的那些物质性质，我们已经给予了说明。在此，可以分条列举如下：

(1) 代谢作用。分子种类的形成和分解必须是互相独立的、自发进行的，即是由正的亲和力所推动的。在任何平衡系统中，都不可能做到这一点，因为其中这两个过程是由微观可逆性互相联系起来的，这种微观可逆性使系统中已出现的所有竞争者产生出某种稳定分布。复杂性，即与时间、空间限度结合中的可选择结构的巨大多重性，简直不可能有这种平衡，而却要求平稳地分解和生成新的结构。选择只有对中间状态才起作用，中间状态从高能前体生成，又将分解为一些低能废物。为此目的系统所应具备的利用这种自由能和物质的能力，称作代谢作用。薛定谔曾第一个清楚地认识到下述必要性，即通过平稳地抵偿熵的产生来保持系统充分地远离平衡态。

(2) 自复制。竞争中的分子结构需有一种内在的能力以指导它们自身的合成。能够表明，这种内在的自催化功能对于任何选择机制都是必要的，这种选择机制包括当出现新的有利突

变体的单独一个拷贝时，即会失去稳定性。而且，对于保持系统中迄今积累起的信息，自复制也是必不可少的。否则，作为条件(1)和(3)的必要前提的平稳分解，将会导致信息的完全丧失。

(3) 突变性。热噪声限制了处于一定温度的自复制过程的精确度。如果复制很快，或更精确地说，如果复制中的每一基本步骤所需的相互作用能并不比热能高多少，那么情况就更是这样。因此，突变性在物理学上总是联系着自复制性，但是它也是进化(在逻辑上)所要求的。复制误差是新信息的主要来源。我们还将见到，对于突变率，存在着某种阈限关系，处于阈值时的进化是最快的，但是不能超过这个阈值，除非在进化过程中已积累起来的所有信息被丢失了。

在实质上是无止境的进化过程中，只有那些满足这3个前提条件的大分子系统，才适合于充当信息载体。相应种类的大分子的全体成员必须具有那些性质，这些成员系指全部可能的选择或给定结构的突变体，而且，它们必须在大的浓度范围内保持有效，也就是从单个拷贝到宏观上可检测丰度的浓度范围内保持有效。"可现实化的前提条件"排除了组分和结构十分复杂的系统，在这种系统中，上述特征将是分子相互作用的某种特殊巧合的结果，而不是物理学相互作用的某个普遍原理的结果。作为一个例子，考虑把核酸与蛋白质加以比较。核酸中的复制是一种一般的性质，其基础是与4种碱基间的独特的互补性相联系的物理力。另一方面，蛋白质的功能容量大得多，其中包括指令性质和复制性质。不过，每一种功能都是多肽链的某种非常特殊的折叠结果，而不能一般地看作各类蛋白质的属性。甚至只要出现一次突变，它也许就会完全丧失。

物质系统，为了能够适合于有选择地自组织，必须具有一

些物理性质，从而可以代谢，即从高能反应物转变为低能产物，以及（"有噪声的"）自复制。这些前提都是必不可少的。还可证明，在适当的外部条件下，对于选择行为和进化行为，它们也是充分的。

1.3.3 选择的动力学

与上述必要前提相一致的最简单系统，可由如下形式的微分方程系统来描述（$\dot{x}=dx/dt$，t=时间）：

$$\dot{x}=(A_iQ_i-D_i)x_i+\sum_{k\neq i}w_{ik}x_k+\phi_i \tag{1}$$

式中下标 i 属于所有可区分的自复制分子单元，从而也标征了它们的特殊的（遗传）信息。x_i代表相应的群体变量（或浓度）。其他参数的物理意义，将在对此方程的讨论中得到阐明。这组方程首先涉及到那些自复制单元 i，它们出现在所考虑的样本中，可从 1 计数到N。这可以被推广而包括所有可能的突变体，其中的一部分突变体在进化期间出现。

在描述开放系统的这些方程中，代谢由分子种类的自发生成（$A_iQ_ix_i$）和分解（D_ix_i）来反映。"自发"意味着两种反应以正的亲和力进行，因此不是相互可逆的。A_i项中总是包含（λ类）高能原料浓度的某种计量函数$f_i(m_1, m_2, \ldots, m_\lambda)$，这种原料是合成分子的种类$i$所要求的，其精确形式则取决于特定的反应机制。这种高能原料必须由流入的物质流平稳地提供，尤如由相应的物质流（ϕ_i）移走反应产物。对于一个自发的分解，D_i项线性地与x_i相联系，反映了普通的一级速率定律。在更复杂的系统中，如果相应的反应是酶促反应，或者如果在反应之间有更多的耦合，那么A_i和D_i都可能包括更多的浓度函数。

自复制是第2个前提，在生成速率项的 x_i-相关中出现。直接的线性相关只代表了内在自催化的最简单形式。其他诸如互补指令或循环催化这种更复杂的、但仍然是线性的机制，可以用类似的方式加以处理。另一方面，非线性自催化是本文的主要对象。

突变性由品质因子Q_i来反映，Q_i可取0至1之间的任何值。这个因子是指发生在一定模板 i 上的复制分数，该复制导致了 i 的准确拷贝。当然，有一个互补项$A_i(1-Q_i)$ x_i与模板i的不精确复制有关。这意味着产生出大量的各种各样"错误拷贝"，它们在大多数情况下都与种类i有密切联系。于是，i 的错误拷贝产生便会在它的每一个"亲属" K 的速率方程相应项中表现出来。相应地，由于复制中的误差，拷贝 i 亦会接受到那些亲属的贡献，这是由它们复制中的误差导致的，这由求和项$\sum_{i \neq k} w_{ik} x_k$对此加以考虑。个体突变速率参数 w_{ki}与复制速率参数 A_iQ_i 相比，往往都比较小，而且，A_iQ_i越是小，则与"亲属"k的关系越是疏远。如果借助下标i和k(取值从1到 N)表示出其中所有物种及其可能的突变体，那么对于错误拷贝，如下守恒关系成立：

$$\sum_i A_i(1-Q_i)x_i = \sum_i \sum_{k \neq i} w_{ik} x_k \tag{2}$$

最后，个体流或运输项ϕ_i描述了任一种类i 的供给或移走，但不是由化学反应供给或移走。这是代谢转变(参见前面)所要求的。在大多数情况下，每一种类对总流ϕ_i的贡献都与其存在成比例：

$$\phi_i = \phi_t \frac{x_i}{\sum_k x_k} \tag{3}$$

在进化实验中，可以通过调节总流，以提供可再现的综合条件，例如恒总群体密度：

$$\sum_k x_k = \text{恒量} = C_n \tag{4}$$

在这种情况下，必须平稳地对流ϕ_t加以调整，以抵偿过量的总产生，即

$$\phi_t = \sum_k A_k x_k - \sum_k D_k x_k = \sum_k E_k x_k \tag{5}$$

我们把式中的$E_i = A_i - D_i$称作模板 i 的"超额生产力"。注意，误差产生并不在这个和中作为守恒关系(2)的结果而明显反映出来。

另外，如果也调整每种高能原料的流量，使 λ 类中每一类都处于恒定的缓冲水平 $(m_1, m_2, \ldots, m_\lambda)$，那么速率参数 A_i 中出现的计量函数 $f_i(m_1, m_2, \ldots, m_\lambda)$则是常量，对此亦就无详细说明的必要。我们把这种

约束称作"恒总组织"，在此约束中，通过控制流量，把所有未组织的以及全部已组织的原料调整到某种恒定的水平。它通常被保留在进化实验中，例如在流动反应器或者通常在某个系列转移实验中。另外，也可选用"恒流"这种直接约束。这时，浓度水平是变量，可调节到给定的流入流量和流出流量的转变点。两种约束都将使得系统趋近某个具有明显选择行为的定态。定量的结果可揭示出两种约束下的差异，但是它们的定性行为原来十分相似。所以，在此只考虑两种限制情况之一就可以了。在自然界中遇到的约束条件可能会随时间变化，所以往往并不相应于其中的任何一种简单的极端情况，这正象天气条件通常并不类似于简单的热力学约束条件(例如恒压、恒温等)。不过，基本的自然选择原理只有在受控和可重复实验条件下才能加以研究。

在恒总组织约束下，速率方程(1)和辅助条件(2)至(5)结合，得到

$$\dot{x}_i = [W_{ii} - \bar{E}(t)]x_i + \sum_{k \neq i} w_{ik} x_k \tag{6}$$

式中

$$W_{ii} = A_i Q_i - D_i \tag{7}$$

此项可以称作(内在的)选择价值，且

$$\bar{E}(t) = \sum_k E_k x_k / \sum_k x_k \tag{8}$$

即平均超额生产力，它是时间的函数。只有当群体变量$x_k(t)$变得稳定时，$\bar{E}(t)$才会取得某个恒定的稳态值，因为这个值取决于突变体谱系的群体分布，所以它是亚稳的。对于恒定的(即不随时间变化的)W_{ii}和w_{ik}值，可以类似地求解非线性微分方程组(6)。关于选择问题的近似解，在较早的论文中已有报道。近年来，由汤普森和麦克布赖德求出了精确解，琼斯、恩斯和兰内卡也分别独立地得到了精确解。由二级摄动理论从精确解得出的显解表达式，与先前报道的近似解是一致的①。以下的讨论以琼斯等人给出的精确解为基础，这个解对于选择问题提供了一个美妙的定量表达式。

1.3.4 "拟种"概念②

由于存在着耦合，所以单个物种并不是独立的实体。总群

体数守恒迫使所有物种进入相互竞争之中，但是突变仍然容许某种协同作用，尤其是密切联系的物种之间（即在具有非零的w_{ik}和w_{ki}项的物种i和K之间）有协同作用。

所以，让我们以如下方式重组我们的系统。我们不是把总群体分解为N个物种，而是定义一个新的N个拟种的集合，其中群体变量y_i是原先的群体变量x_k的线性组合，当然，由此总和是不变的：

$$\sum_{k=1}^{N} x_k = \sum_{k=1}^{N} y_k \qquad (9)$$

由微分方程(6)的结构，提示了如何实现这种新的分解。它实际上相当于对此坐标系进行仿射变换，这在线性微分方程理论中是众所周知的。所获得的关于变换后的群体变量y_i的新的方程集记作：

$$\dot{y} = [\lambda_i - \bar{E}(t)]y_i \qquad (10)$$

把这个程序施用于非线性方程(6)是可能的，因为引起非线性项即(8)式的$\bar{E}(t)$在变换中保持不变，现在可以表示为λ_i的平均：

---

① 琼斯等人指出，在趋近稳态时，因为$W_{mm}-\bar{E}(t)$变得非常小，故忽略逆流项$\sum_{k\neq i} w_{ik}x_k$的做法是无效的。他们指的是，为了展示出典型选择的解的本性而慎重地忽略掉了任何误差率(即$Q=1$)。不过，他们没有注意到我们的详细说明：这种假设只近似地适用于确立了选择优势的占支配地位的种类，而突变细胞的存在只归因于出现了w_{ik}项。先前所获得的近似解的确与应用摄动理论得到的精确解在定量上是一致的。

另一方面，我们应当指出我们十分欣赏由汤普森和麦克布赖德以及琼斯等人获得的精确解，它们大大有助于描绘一幅一致的拟种图象。

② 刘为民对一些数学推导有异议，从而怀疑某些推导结果，见：刘为民，"关于Eigen进化方程的注记"，《科学通报》，22(1981)：1397—1399。——译者注

$$\bar{E}(t) = \sum_k \lambda_k y_k / \sum_k y_k \qquad (11)$$

式中 λ_i 是此线性动力系统的本征值。它们以及将 x_i 同 y_i 联系起来的本征向量，都可以从由系数 W_{ii} 和 w_{ik} 构成的矩阵获得。

系统(10)的解的物理意义是明显的。任一(由本征值λ_i和群体变量 y_i 表征的)拟种，若其λ_i值低于由平均值$\overline{E}(t)$代表的阈值，则它将消亡。(它的速率为负!)相应地，λ_i值大于阈值的任何物种，都将生长。于是，阈值$\overline{E}(t)$是时间的函数，依据(11)式，它将增加；而且，若系统越是有利于具有大的本征值的拟种，则它随时间增加得越快。这将持续下去，直到达到某个稳恒状态：

$$\overline{E}(t) \longrightarrow \lambda_{max} \tag{12}$$

这就是说，平均生产力将增加，直到它等于最大的本征值为止。由此，除了那些具有最大本征值的拟种以外，其余的拟种都将消亡。它们的群体变量已变为零。

因此，达尔文选择和进化能够用极值原理来标征。它定义了在指定的选择约束下自复制体的一类行为。

例如，这样一种过程可以在模拟系统中看到，这代表在孤立的条件下物质系统的一类基本行为，并且是由一般的极值原理来标征的。极值原理与普里戈金和格兰斯多夫的稳定性标准相联系。作为一个最优化原理，对于某些种类的非线性动力学系统也是成立的。而且，准线性系统的解的有效性并不局限于稳态的邻域。

拟种的物理意义是什么？

在生物学中，一个物种即是以一定的表现型行为作为特征的一类个体。在基因型层次上，某特定物种的个体可以有些差异，然而，所有个体都是由结构上非常一致的DNA链所代表

的。它们个体上的区别，就是它们的核苷酸序列上的差异。在此对作为复制单元的这类分子的讨论中，我们也正好是运用它们在序列上的这些差异去定义(分子的)物种。当然，这种差异也是由诸如复制速率、生命期、误差率等不同的表现型性质来表现的。

不过，单个(分子的)物种不是真正的选择目标。(10)式告诉我们，真正的选择目标倒是拟种——即通过选择而出现的、有确定概率分布的物种的有组织的组合。这样，在所有其他的分布中它被选择了。在选择约束下，除了一种拟种以外，其余的全部拟种的确都将消失。拟种与所谓的一个群体的“野生型”密切相联系。

野生型往往被当作标准基因型，它代表突变体分布之中的最佳适应表现型。有可能测定噬菌体基因组的一条独特的序列，这一事实支持了这种关于标准拷贝的一种显性表现的观点。然而，(在魏斯曼实验室中)对噬菌体$Q\beta$的野生型分布进行了更为深入的考察，结果清楚地表明，只有一小部分序列实际上与划归在野生型中的序列完全一致，而其大部分代表着单个或多个错误拷贝的某种分布。只有其平均与野生型序列相似。换言之，标准拷贝也许只出现在小于(有时远小于)总群体的百分之几的范围内。不过，尽管群体的主要部分由非标准型构成，但是(与标准拷贝比较而言)在此种分布中的每一个突变体都出现在非常小的范围内。于是，在检出限度内，总分布揭示了某种平均序列，它与标准拷贝完全一致，从而定义了野生型。前面以精确术语引入的拟种，代表了以一种(或多种)平均序列为特征的有组织的分布。在表1-2中给出了典型的分布例子(与RNA噬菌体$Q\beta$有关)。只有当完全类似于标准拷贝的拷贝真正是优势拷贝时，即只有这种拷贝在此分布中具最大选择价值时，才会出

现这种独特的(平均)序列。W_{ii} 非常接近最大价值的突变体，平均说来将以相应的高丰度出现(参见表1-2)。在某些位置上，它们将使得野生型序列有些模糊不清。如果两种关系密切的突变体实际上(几乎)具有相同的选择价值，那么两者可同时出现在

表1.2

在野生型分布中的标准序列的丰度，由它的品质函数 Q_m 和它的优势 σ_m 来确定。给定个数的核苷酸 γ_m，品质函数可以从涉及到特定的翻读出机制的核苷酸的平均字符品质$\bar{q}_m$计算出来。$\bar{q}_m$和 σ_m两者还决定了核苷酸的最大数目 γ_{max}，标准序列绝不可能超过这个数，否则拟种分布将变得不稳定。这些数据是关于由 4500 个核苷酸构成的 *RNA* 序列(噬菌体 $Q\beta$)的数据。

表 a 的每个右上三角形中的值，显示了按照(18)和(25) 式的野生型分布中标准序列的相对丰度(以百分数表示)。负值意味着该分布是不稳定的。每个左下三角中则显示了给定$\bar{q}_m$和σ_m的阈值数γ_{max}。这些数据清楚地揭示了γ_{max}对于参数$\bar{q}_m$的灵敏性。对于适应得很好的物种，$1-\bar{q}_m$应该稍大于 $1/\gamma_m$ (例如，对于要求 $\sigma_m \geqslant 10$ 的 $\gamma_m=4500$ 的核苷酸，$1/\gamma_m=1-\bar{q}_m=0.0005$)。

表a

γ_{max}		正确拷贝			
		$\bar{q}_m$平均字符复制品质			
		0.9980	0.9990	0.9995	0.9998
优	2	不稳(<0) 346	不稳(<0) 693	不稳(<0) 1386	不稳(<0) 3466
势	20	不稳(<0) 1498	不稳(<0) 2996	5.8% 5991	37.6% 14979
σ_m	200	不稳(<0) 2649	0.6% 5298	10% 10597	40% 26492

表b表示一个更为现实的拟种分布例子，其中个体数为1×10^9。假若对每一个正确的核苷酸(A、U、G 或 C)，有3种不正确的选择，那么4500个核苷酸的序列就会有13500 种单错突变体。不过，*RNA* 复制酶的实验表明，嘌呤→嘌呤和嘧啶→嘧啶的取代，远比任何嘌呤↔嘧啶交叉型取代频繁得多。因此，为了更现实一些，我们已假定，任一位置只有一种不正确的替换。相应地，任一k错误拷贝的多重性正好是$\binom{\nu}{k}$。所有 4500 种不同的单误差突变体，已按其简并的(平均)选择价值划分为 5 类。

表b

突变体分类	每类的简并性	假设的相对选择价值 W_{kk}/W_{mm}	个体突变体的群体数 (×简并数)
无误差	1	1	$8.9\times10^7(\times1)$
单误差			
$M1_a$	1	0.99	$4\times10^6(\times1)$
$M1_b$	4	0.9	$5\times10^5(\times4)$
$M1_c$	495	0.3	$6.3\times10^4(\times495)$
$M1_d$	2000	0.1	$4.9\times10^4(\times2000)$
$M1_e$	2000	~0	$4.3\times10^4(\times2000)$
$\sum M1$	4500		$2.2\times10^8(\times1)$
复误差			
M_k	$\sim10^7$	~0	$<30(\times10^7)$
$\sum M_{k>1}$	$\sim2^{4500}$	~0	$6.88\times10^8(\times1)$

$M1a$ 类中有一种突变体，它极其类似于标准序列，它的选择价值(W_{kk})仅有 1％的差异。M1b 类中包括4种简并突变体，其选择价值与标准选择价值有 10％ 以内的差异，而M1c类中的495种简并突变体，其W_{kk}为 W_{mm} 的 30％。2000种突变体的$M1d$类，其 W_{kk}低了一个数量级，并且相同数量的突变体根本不会成活，即它不可能以可与标准速率相比的速率进行复制。而且还规定所有具有两处以上错误的拷贝，其值$W_{kk}\ll W_{mm}$。尽管这也许不是一种现实的假定，但是对于个体序列的群体数它不会有严重影响，由于不同错误拷贝的极大多重性，个体序列为数极小。暂不考虑这种事实，在这个例子中，多错拷贝的总和为最大的一群，其次才是单错拷贝的总和。另一方面，在拟种分布中，标准序列的个体数最多。

另一种计算中，单错突变体1a的相对选择价值已从0.99提高到 0.9995。而总分布只发生轻微的变化，特定突变体1a的群体数提高到等于标准序列的值(即两

者的结果都是8.4×10^7)。这个例子表明，在(18)和(19)式背后近似的程度，方程(18)与(19)要求$w_{kk}\ll W_{kk}-W_{mm}$更为严格的数值评价得出，突变体 1a 的群体数总计仅为标准序列的群体数的 60%。哪怕选择价值只有很小的差异，标准序列显然仍保持为优势物种。只有某种单错突变体在 $W_{mm}-W_{kk}\ll 1/\gamma m$ 的限度内类似于标准序列，它才可以被看作是简并的，从而是难以区分的个体。

具相等的统计权重的拟种中。对于两种互相类似的突变体来说，W_{ii}值要接近到什么程度才变得选择上不可区分，这取决于它们的"亲和度"。这于远亲，这种相关性必须比单错拷贝精确得多。因此，可以定量地定义一类特殊的"可逆中性"突变体。当然还有更大的另一类中性突变体，它们属不同的拟种，其本征值λ_i是简并的。绝大多数这种中性突变体在出现后即会消亡，但其中也有很小的一部分可以通过群体而扩散，进而与先前确立的拟种共存，或者是将它们取而代之。只有以随机理论为基础，才能理解中性突变体的这种扩散传播(参见后面)。

1.3.5 现实的近似

借助于摄动理论，可以清楚地表示被选择的拟种的本征值。二级摄动理论的结果，类似于前面报道的 W_{max} 的表达式：

$$\lambda_{max}\approx W_{mm}+\sum_{k\neq m}\frac{w_{mk}w_{km}}{W_{mm}-W_{kk}} \tag{18}$$

式中下标m指的是由最大选择价值所区分的(分子的)物种。若没有其他的W_{kk}十分接近这个值，且优势拷贝m可被看作野生型的代表时，这种近似才成立。表 1-2 表明，这种近似对于任何具有现实重要性的系统，确是非常有效。信息量越大，则个体的w值越小。从而这个近似揭示了一个非常重要的事实：(在约束条件下)对于远亲来说，选择是极为激烈的(w_{mk}和w_{km}值越小，则W_{kk}可以越是接近W_{mm}，而对m无任何限制)。不过，对亲缘非常密切的拷贝，选择则是平缓的.如果它们的选择价值W_{kk}比W_{mm}小

得多(或者甚至为零)，这种分布中总是会出现上述情形的。若(13)式中的求和项在数值上可以忽略的话(参见表1-2中的值)，则极值原理[(12)式]可以表示为：

$$W_{mm}>\overline{E}_{k\neq m} \tag{14}$$

或者

$$Q_m>\sigma_m^{-1} \tag{15}$$

而

$$\overline{E}_{k\neq m}=\sum_{k\neq m}E_k x_k/\sum_{k\neq m}x_k \tag{16}$$

代表了所有被选择的野生型m的全部竞争者的平均生产力，且

$$\sigma_m=\frac{A_m}{D_m+\overline{E}_{k\neq m}} \tag{17}$$

是优势种的优势参数。

运用同样的近似，可计算相对静止的群体数，对优势拷贝得到：

$$\bar{x}_m/\sum_{k=1}^{N}\bar{x}_k=\frac{W_{mm}-\overline{E}_{k\neq m}}{E_m-\overline{E}_{k\neq m}}=\frac{Q_m-\sigma_m^{-1}}{1-\sigma_m^{-1}} \tag{18}$$

对于单错拷贝得到：

$$\bar{x}_{1k}/\bar{x}_m=\frac{w_{km}}{W_{mm}-W_{kk}} \tag{19}$$

只要$w_{km}\ll W_{mm}-W_{kk}$，它便成立(参见表1-2)。

运用不论何种形式表示的λ_{max}可获得更高级的近似，前面的(16)～(18)式应作相应的改变。

以这些近似为基础，我们可以定量地表示大分子系统的达尔文行为的特性。(13)式指出选择的动力学在何种程度上取决于优势(标准)种(m)的个性，而(18)和(19)式表明了标准物种及其突变体所表现的相对权重(也参见表1.2)。令人惊奇的是，尽管标准物种的物理学参数几乎完全决定了此分布的动力学行为，但在野生型分布中只出现很小的一部分标准物种。这提供了拟种的很大适应和进化的灵活性，从而能对环境变化作出迅

速反应。

只有出现了两种或更多(几乎正好是)中性突变体的优势种时(参见表1.2)，这种近似才失效。不过，这些可逆的中性突变体可以结合成为某种选择上不可区分的物种的亚类，从而将按照上述的关系决定类似于单一优势种的动力学行为。注意到如下事实是有趣的：在拟种之中不存在对中性突变体的选择(“可逆”被定义为具有充分大的w_{ik}和w_{ki}，以保证某种可复制的样本)。这些作为拟种的一部分的可逆中性突变体必须与无关的中性突变体加以区分[w_{ik}和w_{ki}太小，据(18)式，这种突变体难以提供某种可复制的样本]。那些无关的中性突变体仍然能够在很小的范围内作为随机涨落的结果而共存。随机处理表明，在那些中性突变体中间的竞争，是一种随机漂移现象，它导致不明确的“生存者生存”的消失，也导致了新的突变体的某种成长和扩散；这种现象被遗传学家称作非达尔文行为（即没有选择优势而生存)。应该认识到，无关的中性突变体的这种随机行为，尽管是达尔文及其追随者所不希望的，但是它并不与导致决定论的达尔文行为的那些性质发生矛盾。如果正确地把达尔文原理解释为可推导的物理学原理，从而将其应用于拟种概念。那末甚至可逆中性突变体的选择也将符合达尔文原理。

1.3.6 推广

由(2)和(6)式定量地表示达尔文系统时，我们作了一些与结构前提和外部约束相联系的特殊假定。在这一节，我们希望能发现，在不失达尔文行为特征的情况下，能把那些假定推广到什么程度。

1)速率项。自催化生成以及一级分解的线性速率项，都可以用更一般的表达式来代替。对于酶机制，最普通的是米凯利斯一门顿形式：

$$速率\sim\frac{x_i}{1+a_i x_i}或\frac{x_i}{1+\sum_k a_k x_k}$$

它替代了简单的x_i相关。业已表明，对于这种自催化机制，选择将对于小

的群体保持有效，因之在此范围内它对于选择是关键性的。饱和不会阻止优势突变体的成长，但是由于未约束的机制从指数生长变为线性生长，从而允许某些共存。

一般而言，如果反应级数由 x_i^k 项所定义，那么达尔文行为可能有利于指数：

$0<k\leqslant 1$，

对于 $k=0$，共存会导致一个有限生长系统。在恒流约束下，对于自复制物种可能发生这种情况，如果它们的生成速率受到以恒定速率供给高能原料的限制，那末，以不同的（互相独立的）原料源为生的物种，互相之间将不会竞争。独立的来源则为共存提供了"小生境"。人们往往把多种多样的不同物种的存在归因于这样的或相似的处境，其结果与达尔文理论的看法完全一致。

对于指数 $k>1$ 的行为，在本文的第二部分中进行了更为详尽的分析。这种行为导致了激烈的选择，其中某些结果（关于后裔方面的）尤其与达尔文的观点不相容。

(2) 自催化。依照我们在第2节的讨论，直接的自复制不过是整类线性自催化机制中的最简单例子。通过循环催化过程，一般会影响自复制。例如，单链 RNA 噬菌体的复制受到互补的两条链的相互指令。这种2元催化循环的速率方程产生出关于优势种的两个本征值：

$$\lambda_{1,2}=-\frac{D_{+}+D_{-}}{2}\pm\sqrt{A_{+}A_{-}Q_{+}Q_{-}+\frac{1}{4}(D_{+}-D_{-})^{2}} \qquad (20)$$

这些表达式以类似于诸如(13)式那样忽略掉突变项的近似为基础。如果其中的一个本征值为正（即 $A_{+}A_{-}Q_{+}Q_{-}>D_{+}D_{-}$），就可用它代替此自复制单元的 W_{mm}。在此，两条链的动力学参数以相等的权重（几何平均）对选择价值作出贡献。我们必须对它们进行优化处理，以产生出优化行为。RNA 链的表现型性质无论在何处都是重要的，可以由结构对称很合适地达到等价（参见tRNA，Qβ RNA的中变体）。第二个总是为负的本征值涉及到正链和负链之间的某种"平衡"，于是它们的浓度呈现一固定的变化。无论何时正链和负链达到这种平衡，它们就象一个复制和竞争的单元一样起作用。

一般处理业已表明，催化循环类似于简单自我指令系统的结果。而一个n元循环又是由相应于单自复制单元的选择价值的一个正本征值标征的。所有成员的催化性质在最简单的情况下，以其AQ值的几何平均的形式对这个本征值作出贡献，因之要求此循环中的所有成员都具有确定的AQ值。而且，一个n元循环，其特征是有$(n-1)$个负的本征值，它们代表了关于循环中全体成员的浓度比的内平衡。

(3) 突变。特别是在进化的初期阶段，突变的主要来源是错误复制，即在复制过程中包含了带有非互补碱基的核苷酸。用$Q\beta$-噬菌体进行的试验表明，某一给定的嘌呤或嘧啶被其同源体取代的误差率，大大不同于嘧啶交换嘌呤或嘌呤交换嘧啶的误差率。在运用参数Q和w进行的形式处理中，对诸如点突变和由于缺失或插入形成的移码突变等突变现象，不必加以区分。若要研究突变体的功能性质，则作出那些区分是重要的。对于诸如复制中的错误阅读、化学诱导变化或辐射损害等所引起的突变，此形式表达式仍然不变。把某些突变与分解项更紧密地联系起来可能是必要的(参见下面)，不过，这对于方程的形式结构毫无影响。

(4) 分解。作为外部影响(例如辐射)结果的突变，应该被归属到分解项中。一般地说，各个i的分解，可以导致仅由i的碎片组成的另一类物种k。如果恰当地考虑到相应的守恒关系，那么那些过程也不改变微分方程的形式结构。

(5) 外部约束。选择方程的显解形式，依赖于附加的约束条件。我们已经详细讨论了恒总组织的情况。同样，尽管有定量的差异，但是获得了恒流约束下的结果。当然，外部调节的参数可以包含有时间变化(例如任何类型的周期变化)。这可以导致力学优势，但不会根本改变达尔文行为的基本前提和结果。极值原理(12)式则取得一般形式：

$$\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau}\bar{E}(t)\,dt=\lambda_m \tag{21}$$

琼斯等人已对极值原理作了进一步推广。

关于某种拟种相对于它的竞争者的选择，也可以在生长系统中加以考虑。它将表明，一种简单的规范化程序是适用的，从而可对非稳态系统进行一般的处理。

(6) 随机处理。最后的推广更为基本。决定论的速率方程一般描述由大量个体构成的系综的平均行为。不过，基本过程却只能反应概率来表示。我与温克勒-奥斯瓦瑟共同发展起来的博奕模型清楚地表明，有 3 类基本行为可以用随机理论进行处理：

a) 涨落的内部自控制在稳定的稳态附近，尤其是热力学平衡态中找到；

b) 涨落的自放大描述一种不稳定性；

c) 产生随机漂移行为的涨落是无关紧要的。

在第 1 种情况下，涨落只对于小的群体数是重要的，它只不过意味着任何瞬时微观态的某种不确定性。对于可由实验检验的宏观态，它们产生出在可说明的平均涨落限度以内的期望值。

在第 2 种情况下，决定论的行为被限制在对一定的涨落作出反响。换言之，如果出现了某种涨落，那么这种"如果一那么"决定性能预见将发生什么，预见的准确性还将随着涨落的程度而增加。不过，涨落本身的出现是不确定的，这种微观的不确定性通过最终的决定论的放大过程在宏观上映射出来。这种情况与达尔文系统特别有关。大多数突变是第 1 类涨落，即它们没有任何选择优势，并且不危及野生型的稳定性。它们出现以后肯定会停息下来，如同处于平衡态。然而，仍然存在一些会带来选择优势的突变，它们具有放大自身的趋势，从而引起某种不稳定。它们是否能成功地取得支配地位，要取决于它们的选择优势的大小和涨落的程度。单独一个拷贝，在它被复制之前，尤其是如果它的W值只略大于$\bar{E}(t)$的话，它面临消亡的机会仍然是很大的。随机理论表明，对于小的优势，即对于$(W_{m+1}-W_m)\ll W_m$，在生长概率变得大于$1-e^{-1}$之前，涨落就必须达到一定的程度，也就是必须有对应于$W_m/(W_{m+1}-W_m)$大小的拷贝数。这就等于说，只有具有明显优势的突变体，才将决定性地影响进化行为。接近中性的突变体行为几乎随机地类似于真正的中性突变体，所以，它们涉及到第 3 类博奕，这类博奕类似于随机漂移行为。

频繁出现的中性突变体，当然是给定拟种的一部分，从而由于一定的突变速率而多少有点稳定。因此，它们利用了作为第 1 类特征的某种涨落

反应。因为这种耦合是相当弱的，在中性亲缘拷贝的亲属代表中可能有相当大的涨落。另一方面，无关的（即非常稀少的）中性突变体可被看作不同的拟种，其本征值的大小与野生型的相同。大部分中性拟种都将消亡，但是，它们一旦成长起来，就可以继续生存下去，甚至会取代先前的野生型（“生存者生存”）。只有从随机理论才能导出这种行为。对孟德尔群体中基因扩散已经做了相应计算，其中尤其以木村资生及其学派的工作最为显著。

总之，进化就其进步特征而论是一种决定论的过程。在此总是存在着有利于野生型与优势较小的突变体之间的竞争，总是共存着中性的或接近中性的密切相关的突变体，以及总是会成长起新的有明显优势的拟种。不过，就突变体出现的时间顺序而论，以及就由无关的中性突变体引起的基因漂移而论，进化不是决定论的。这些中性拟种中，实际上只有一小部分能够成长起来。所以，通过稀有的中性突变引起的进化，在进化的后阶段比起在其早期阶段更为重要，在这早期阶段仍然可能有多种有利的变化，从而以比较高的频率出现。

1.3.7 拟种的信息量

在前面对分子进化的探讨中，我们还没有明确地讨论遗传信息概念。我们已把（分子）物种定义为具有一定信息量的复制单元，其信息量由分子符号的特定排列来表示。我们还注意到在各种符号排列之间的相似性关系，这种排列导出了拟种概念。在推导选择和进化的标准时，它充分表明，正如在群体遗传学中一样，须注意到基因型信息的个体差异，并把它们与由选择价值W_{ii}表示的特殊的动力学性质相联系。

另一方面，诸如“在一定的拟种中能够积累多少信息?”或者“可复制的限度在哪里以及从而拟种的进化能力的限度在哪里?”此类问题，在上面的讨论中并未作出回答。因此，我们现在集中谈谈引入信息概念的问题：

在通信理论中，由γ_k个符号构成信息的信息量可以表示为

$$I_k=\gamma_k i \tag{22}$$

式中i是单个符号的平均信息量。按照香农的看法，i可以与符号的概率分布联系起来：

$$i=-K\sum_j P_j lnP_j \tag{23}$$

其中

$$O<P_j<1,\ \sum P_j=1$$

为得到单位“比特/符号”，常数K通常取$1/ln2$。字母表一般包括若干类符号(例如，对于核酸，$\lambda=4$)。方程(23)的实际应用只限于如下情况：所有符号的先验概率是已知的，信息中的符号数充分大，因而可以运用平均数。为了说明所有的协同效应或影响概率分布的冗余度，可能有必要知道符号结合全部λ^{γ}种可能的选择对象。

在此，我们的兴趣实际不在于一个符号的任何统计的先验概率，而在于一给定符号由遗传机制正确复制的概率，不管在各种组织层次上这种机制有多么复杂。这种概率涉及到信息转移的动力学过程，因此必须由动力学数据从实验上加以确定(只要可能就必须这样做，参见后面)。让我们把这些正确符号复制的概率记作q_j。由γ_i个(分子)符号构成的一条信息正确地被复制的概率或品质因子为：

$$Q_i=\prod_{j=1}^{\gamma_i} q_{ij}=\bar{q}_i^{\gamma_i} \tag{25}$$

即使是符号仅由λ类构成，信息的每一符号的品质因子可能会依赖于它的特定环境，因此，确定几何平均值$\bar{q}$要求考虑到各种协同效应，尤其是与信息“i”相联系的协同效应。不过，对于任何给定的信息i可以确定这个平均值$\bar{q}_i$，结果表明，只要此种信息充分长，则那些平均值相当普遍地适

用于特殊的酶复制机构。而且，单个 q 值一般都接近 1，以致几何均值可用算术均值来代替，即

$$\left(\prod_{l=1}^{\nu_i} q_{l_i}\right)^{1/\nu_i} \approx \frac{\sum_{l=1}^{\nu_i} q_{l_i}}{\nu_i}, \text{ 若}(1-q_{l_i}) \ll 1 \tag{26}$$

(25)式于是成为复制的信息-理论方面，不过，在此 $\bar{q}$ 指的是动态的概率而不是静态的概率。$\bar{q}$ 的数值可能要考虑到符号复制的所有机构的特点，其中包括减少复制过程误差率的任何静态冗余。大自然实际上已经发明了精妙的复制设备，其中包括了从互补碱基识别到复杂的酶检及校对机制。

遗传复制是连续的自我重复的过程，从而不同于信息通过噪声信道的简单转移。在此的每一次转移，所要求的不仅仅是再现信息的意义，它所给出的某种冗余总是使一部分符号被不正确地复制。阻止一系列相继的复制循环中积累更多的错误，也是必要的。换言之，必须能够有利于一部分完全正确的野生型拷贝能与它们的全体错误拷贝进行有利的竞争。只有这样，野生型才能保持稳定的分布。否则，信息(现在是在语义学意义上，即拷贝具有最优 W_{ii} 值)将会慢慢丢掉，直到丧失殆尽。

稳定保持信息的条件，是选择标准(14)式或(15)式。这个标准能够以一般形式表示为：

$$Q_m > Q_{\min} = \sigma_m^{-1} \tag{27}$$

从而可适用于任何复制机制，哪怕 σ_m 不可能以如同对线性机制适用的那种简单形式[参见(17)式]表达出来。如果我们把这个从动力学理论导出的选择标准，与由(25)式表达的信息方面结合起来，我们就获得了关于拟种的最大信息量的重要阈值关系：

$$\nu_{\max} = \frac{\ln \sigma_m}{1-\bar{q}_m} \tag{28}$$

一个自复制单元的分子符号数是受限制的，其极限与每个符号的平均误差率$1-\bar{q}_m$成反比。

还可用另一种方式表示出这种重要的关系。在一条γ_m个字符序列的某个误差期望值$\epsilon_m=\gamma_m(1-\bar{q}_m)$，必须总是保持在某个明确的阈值之下：

$$e^{\epsilon_m}<\sigma_m \tag{29}$$

否则，只要一次错误突变，就会使得进化过程中积累起来的信息丢失掉。表 1-2中 的一些例子表明了与(28) 式的这种联系。此阈值对优势函数的大小并不敏感，但是σ_m必须大于1，即是$\ln\sigma_m>0$，以保证γ_{max}的确定值。实际上(参见后面)，$\ln\sigma_m$通常在1和10之间。对于任何特殊复制机制都能提供的进化潜势，可由关系式(28)进行定量估计。例如，它指出1%的误差率(或说99%的符号复制精确度)，就足以可收集并可复制地保持某种不大于数百符号的信息量(取决于$\ln\sigma_m$的值)。或者说，保持象大肠杆菌大小的基因组的信息量，则要求在10^6至10^7个核苷酸中误差不超过1个。这正是一种可以由实验加以检验的关系，我们在后面也将提呈相应的测量结果。对于其重要性早就被强调的群体遗传学中的所谓“遗传负载”，(28)式也给出了定量的解释。

本节的结果可以小结如下：任何一种信息的选择性积累机制，都涉及到字符数量的上限，这些字符将以某种特定顺序装配起来。如果超过了这个限度，那么，等价于信息的有序，将在相继的复制中溃散。信息稳定性相当于拟种的内稳定性。它尽量以竞争为基础，就象是拟种的选择。然而，两种拟种可以形成共存，而违反阈值关系则将导致遗传信息完全丧失。所以，拟种分布的内稳定性，比起它的“生存斗争”，更是达尔文行为的特殊属性。

1.4 误差阈和进化

1.4.1 误差突变的计算机试验

阈值关系[(28)式]的物理内容，可以通过几种计算机博奕(参见表1.3)来举例说明，目的是要从或多或少的随机字母序列产生某条有意义的信息。为此，把一组N个随机序列输入计算机，并编写程序：

(1)每一序列，当它到了规定的(平均)寿命后，就从计算机的存贮中抹掉，

(2)对每一条以特征速率从存贮中出现的序列进行复制，并且

(3)以每个符号取代的选定平均速率，在复制的拷贝中引入随机误差。

(1)以及(2)的平均速率以如下方式匹配：语句的N个拷贝的稳定显示保留在存贮器中，每条语句都有一定的寿命。因此，任何在博奕期间获得的信息，都只有通过所出现的序列进行忠实复制才能得到保持。另一方面，信息的获得，必定是来自对各种突变体序列按其意义所进行的选择性评价，或最好说是按其与任一意义或多或少密切的关系所进行的选择性评价。这种评价要受到序列所固有的依赖于与意义有关的顺序性质的作用，从而影响到它们的复制和(或)消去的速率，这正如选择价值所表达的。

表1-3

语句的自我纠正是进化博奕的结果，在本表中举例说明这一点。

设定目标语句为：

TAKE ADVANTAGE OF MISTAKE

选用这条语句，是因为它提供了关于进化机制的"选择优势"信息。无论何时在单词之中引入了动能联接（正如第二部分所要实施的），其特定形式就允许某种循环封闭。

运用一种码，其中每一字母(和词距)都由5个一组的二元符号来表示，其信息量为$\gamma_m=125$比特，即可有4×10^{37}种选择。这个数把仅仅是偶然出现的信息排除在外。本表中显示的一定世代的字母序列，已当作存贮在计算机中的序列总群体的代表而予以取样。

起始序列：BAK GEVLNT GUPIF LESTKKM

字符品质$\bar{q}_m$:0.995

每比特选择优势：10

第1代：RAK GEVNNT GUPQF KESTKKM

第5代：NAK AEZ, NS GEPOF MESTMKU

第10代：VAKF ADVINT. GE OF MISDIKE

第16代：TAKE ADVANTAGE OF MISTAKE

(目标达到)

第1个例子表明，在接近临界值 $1-\bar{q}\approx1/\gamma_m$时($\gamma_m=125$，相应于$\bar{q}_m=0.992$)，进化是十分有效的。从某个随机的字母序列出发，对于任何在 0.995和0.990之间的$\bar{q}$值，通常在20(±6)代之内达到目标语句。如果对于一定的世代，比较各种$\bar{q}_m$值的进化进步，这种有效性在接近阈值时甚至更为明显：

起始序列：BAK GEVLNT GUPIF LESTKKM

每比特选择优势：10

$\bar{q}_m$	11代以后的最好序列	错误数
0.999	LAKD AEV, NTAGU AF KISTQKM	9
0.995	TAKE ADVINT GE OF MISTAKE	2
0.990	TAKEBADVINTAGE OF MISXAKE	3
0.985	VATA ADBKMDI DHOD? CSYBKE	18

对于$\gamma_m>\gamma_{max}$，发现了信息溃散有类似行为。在误差率$(1-\bar{q}_m)=1.5\%$时，每比特选择优势2.5相应的σ_m值约为5。在这些条件下，信息不再稳定。不过，对于小的选择价值(如在本例中)，信息的溃散(或积累)是一个相对缓慢的过程。

起始序列：

字符品质$\bar{q}_m$: 0.985

每比特选择优势：2.5

世代数	最好语句	错误数
1	TAKE ADVANTAGE OF MISTAKE	0
5	TAKF ¦DVALTAGE OF MISTAKE	3
10	TALF ADVALTACE OF MISTAKI	5
20	DAKE ADUAVEAGB OF MJUTAKE	6
40	TAKE ADVONTQCU OF MFST¦ME	7
71	TAKEB? VALTAGI LV MIST¦KE	8
对于$\bar{q}_m$=0.97的71代：		
	?AMEBADTIMOACFHQEBA! STBMF	18

把最后两行(都是71代)加以比较表明，在误差率为3%($\bar{q}_m$=0.97)时，溃散快得多。

对另一个(每比特选择优势=10)的例子，在$\bar{q}_m$=0.985时还不会超过阈值；所以，如下面所见，信息是稳定的，此时过程是从正确语句开始的。

起始语句：TAKE ADVANTAGE OF MISTAKE

字符品质$\bar{q}_m$：0.985

每比特选择优势：10

71代 (打印出8条代表语句)	错误数
TAKE ADVANTAGE OF MISTAKE	0
TAKE ADVANTAGIPOF MISTAKE	2
TAKE ADVANTAGE OF MISTAKE	0
TBKE ¦DVANTAGE OF MISTAKE	2
SAKE ADVANTAGE OF MGSTAME	3
TAOE ADVANVAGE OF MISTAKE	2
TAKE ADVAVTAGE OF MISTAKE	1
TAKE .DVANTAGE OF MISTAKE	1

在自然选择中，进化的目标总是这种基因型，它代表了具有最优选择价值的表现型。而表现型影响到评价。表现型是特定个体的物理和化学性质，这些性质在群体内竞争者的平均水

平上，与基因型的复制速率、品质和寿命有关，类似地，在我们的心理记忆中，能够把任一字母序列与联系着其意义的复制价值结合起来。从而运用想象，在博奕中我们可以对任一序列，按其与意义的联系进行评价。当然，计算机只有通过特定的程序才能做到这一点，此种程序中进化的语句要与一条或二条可能的目标语句进行比较。

那么，让我们假定，每一序列以某种速率被复制，此速率依赖于符号的数目，这些符号在它们的位置上与(有意义的)目标语句的符号相一致。运用二元编码，我们或许可以明确表示，每一位向目标接近一点，我们按某个系数提高复制速率；而每一位远离目标一点，则相应地使复制速率减速。

不过，必须预先输入计算机的关于可能的目标语句的信息，除了把某个有用的方案提供给计算机以外，就再没有其他用处了。容易把评价程序做得更为高级，使之更接近我们的思维对“意义”的评价，直到能够完全不必预先设定最终所要达到的特定目标。然而，在我们很好地定义了内在的分子进化评价程序以后(演示*tRNA*分子进化的博奕模型提供了一个例子)，对于例示阈值关系的物理意义来说，这些模拟的机制细节就不那么重要了。依据表1-3，可以把计算机实验的结果总结如下：

对于高品质的符号复制，例如，当100比特的语句的每符号平均误差率$1-\bar{q}\ll 10^{-2}$，即使有大的σ_m值即具有大的选择优势，进化的进步仍是很慢的。如果选取与$1/\nu_m$有同样数量级的平均误差率$(1-\bar{q}_m)$(例如，对于100比特，$\bar{q}\approx 0.99$)，可以取得最大进步。

对于充分大的σ_m值(>3)，则目标语句可在若干世代内获得，此世代数对应于目标序列和起始(或多或少随机的)序列之间进化距离的数量级（例如100代)。不过，一旦超过了$1-\bar{q}=$

$\ln\sigma_m/\gamma_m$的阈值，无论将每个比特的选择优势选得多么大，都不可能获得更多的信息。如果从接近正确的语句出发，序列就会溃散成为一堆字母的随机混合物，而不会进化为无误差拷贝。阈值是很明确的，但是溃散速率在靠近阈值处发生变化。除非σ_m这个参数非常接近于单位值，否则阈值与σ_m值的关系不大。优势σ_m可从相对优势来计算，计算时还需要一些关于误差分布（与相应的最优拷贝有关）的知识。这种分布当然取决于选择优势的大小。计算机实验结果十分类似于预期的误差分布，在接近临界值$1-\bar{q}_m=1/\gamma_m$（即$\ln\sigma_m\approx1$）处，它产生了一个最优拷贝的几乎相同的代表、全部单错拷贝（相对于最优拷贝而言）以及全部多错拷贝（在其分布中又主要是双错拷贝，拷贝数随误差增多而表现强烈减少的趋势）。对于较小的选择优势（例如$W_{mm}-W_{kk}<3$），这种情况向有利于误差拷贝的方向漂移，而不利于（相对的）最优拷贝，在$\ln\sigma_m=1$时，最优拷贝数已不到总数的10%。

1.4.2 *RNA*噬菌体的实验研究

这种博奕，人们如果领会了其结果，它就是相当平常的，它相应地出现在自然界中，决定了在前细胞和细胞的自组织的各种水平上所获得的信息。一个几乎完全类似于上述博奕的实验，已经由魏斯曼及其同事们用噬菌体Qβ实现了。

通过定位诱变，已产生出此种噬菌体基因组的一种误差拷贝。方法是，在试管内合成负链噬菌体RNA，它在5′端的39号位有诱变的碱基模拟物N^4羟基 CMP，以代替原先的核苷酸UMP。以这条链作为模板，运用 Qβ 复制酶，便能获得感染的正链，正链中3′端的40号位相应于负链中5′端的39号位，正链中的这个位置落在外作用子区域——碱基A被G取代了。大肠杆菌原生质

球就会被这种突变正链感染，产生出完整的突变噬菌体微粒，这能从单噬菌体斑回收到。在体内（大肠杆菌由完整的噬菌体微粒的感染）以及在试管内（运用Qβ复制酶对分离RNA链的速率进行研究）的系列试验中，可以测定包括其随体分布的野生型和突变体40的复制速率参数。把指纹图谱和序列分析结合起来，运用于相继的世代，从而揭示出在突变群体中的变化是由于生成了回复突变体。用野生型和突变体的不同起始序列进行的研究揭示，自然选择包括了一种优势个体和突变体分布之间的竞争。定量的估算表明，价值依赖于特定的选择优势以及突变体群体的分布参数。与此特定的突变体相比，野生型显示出某种选择优势：

$W_{野生型}—W_{突变体}\approx 2$到4

同时，估计取代速率约为

$1—q\approx 3\times 10^{4}$

q值以回复突变体形成的速率为基础，因此适用于特定（互补）取代：

分别为G→A　或　C→U

依据(20)式，正链和负链的品质因子对复制精确度具有相等的贡献。因此，G→A和C→U取代也是等价的。它们可能与A→G和U→C取代没有太大的差别，其主要原因在于GU和UG相互作用的变位相似性。

因为复制酶要求模板伸展，以便于与活性部位结合，所以q值不会更多地依赖模板区域的二级或三级结构。在试管内对Qβ-RNA的中变体的研究中，所得到的C→U取代的速率值与上述值相吻合。嘌呤→嘧啶和嘧啶→嘌呤取代似乎很少出现，因此实质上对$\bar{q}$的大小无影响。

测定σ_m要困难一些，因为它依赖于$\bar{E}_{k\neq m}$的大小。首先应该

注意到，修饰外作用子区域对复制速率有相当大的影响，但它不影响噬菌体RNA对任何蛋白质编码。西格尔曼最先强调，噬菌体RNA分子的表现型性质对于复制和选择的机制具有重要意义。上面报道的 σ_m 值，指的是某特定的突变体及其随体。其他突变体也许会以不同方式影响Qβ-RNA 的三级结构，从而表现出不同的复制速率。而且，在内作用子区域的突变也许是致命的，因此对$\bar{E}_{k\neq m}$毫无贡献。如果我们考虑代表了大部分突变的测量值，那么我们所获得对于某个最大信息量的值，只会稍大于Qβ基因组的实际大小，此Qβ基因组中包含约4500 个核苷酸。人们也许会对这种高度吻合有几分怀疑，我们也已提出对此持一定的保留。不过，它们主要涉及到σ_m值，它仅仅作为对数项加入进来。更大的σ_m值仍将对ν_{max}产生出可接受的极限。因此，所获得的这个值也许最终并不远离现实。由魏斯曼及其同事实现的另一组实验表明，野生型分布中存在着相对少的一部分标准噬菌体。这些数据表明，$\sigma_m^{-1} \approx Q_m$，并且实际的核苷酸数的确非常接近阈值$\nu_{max}$[参见(18)式]。

在米尔斯和西格尔曼等人的进化实验中所运用的中变体，仅由218个核苷酸组成，因此不象Qβ-RNA 那样能很好地适应环境变化。当然，它是优化地适应了“标准反应混合物”的特殊环境，这是试管实验中所具有的(这不要求RNA 微粒被感染)。不过，它对环境变化的反应是相当慢的，例如对加入复制抑制剂乙基溴化物的反应。经历了20次转移后获得的突变体，每一个可以有十万倍的放大率，与中变体的野生型相比仅仅有 3 个位置不同，从而在新环境中显示出比较小的选择优势。慢响应的原因在于，平均单字符品质为0.995的218个核苷酸产生出的 Q 值接近1，并导致了野生型序列，它们非常忠实地得到复制，在它们的误差分布中始终只有少量突变体($\lesssim 10\%$)。

从理论方面看，这些研究中令人注目的结果并不是阈值关系(如不等式)得到满足。因为导出此关系时，是以相当一般的逻辑推论为基础的，所以，任何大的偏差都将意味着我们对达尔文系统的理解有重大概念错误。任何此种偏差，其原因不是无足轻重的，因为我们对于象噬菌体$Q\beta$这样一种"清晰"系统的分子自复制机制已有了很好的了解。真正惊人的结果是，实际的ν值不仅保持在ν_{max}之下，而且事实上非常接近阈值。核苷酸数可能已被符号品质$\bar{q}$以外的一些因子轻而易举地加以限制，由此产生出远低于阈值的ν值，这个阈值容许重复积累信息。

所以，我们只好得出结论，这些RNA噬菌体在其进化期间，的确致力于积累起尽可能多的信息，并在其基因组中利用了外作用子的、但表现型上是活跃的区域。这种事实并不排除：在其他(例如人为的)条件下，象上述的中变体那种很小的RNA分子能在竞争中获胜，或者，在不同的自然环境下，存在着也是很小的能发育生长的噬菌体。更为重要的是要认识到，自然界不存在基因组中含有多于约10000核苷酸的(单链)RNA噬菌体。这意味着，RNA复制的酶机制，尤其是对于碱基A和G或U和C的识别力，已达到了它的最优状况，不可能再进一步有所改进了。对于任何单链RNA，要能够可复制地保持多于与1000到10000个核苷酸的数量级相当的信息量(精确的数字取决于σ_m值)，是不可能的。

当然，按照化学标准，还可以存在更大的分子，但是它们没有进化价值。而且，正如(20)式表明的，正链和负链都必须满足有选择的保持信息的条件，尽管其中只有一条链需要携带由宿主机制翻译的遗传信息。这些结论只可应用于RNA分子，它在复制时期充当单链模板，对此已有人提出了如图1-11所示的

机制。

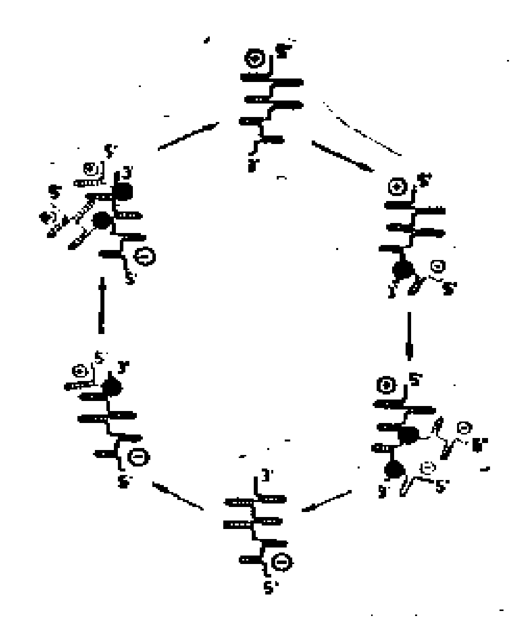


图 1-11 RNA噬菌体中的复制循环 导致单链RNA单元，它们具有被用作表现型目标的高度专一的二级和三级结构。在正、负链之间形成完全的或部分的复式(即分别是所谓的霍夫施奈德或富兰克林配对)，受到新合成链的紧靠着内部折叠的阻止。西格尔曼及其同事运用噬菌体 Qβ 的一种中变体，能够证明由形成复式导致的抑制效应。单链复制的基础在于，复制酶与正链负链二者之间有某种有效的相互作用，要求三级结构中那些对表现型是重要的区域有某种对称性。

1.4.3 DNA复制

对于双链分子，特别是DNA分子，情况大不一样。它们一般以双链的单元进行复制，也就是说，至少在唯象的意义上(哪怕传递指令的基础也是核苷酸的互补性)，它们可被看作真正的自复制。

让我们把已认识到的这种双链 DNA 分子复制的知识简要总结如下(参见图1-12)：

(1) 复制是一种半保守过程。在复制中，DNA 螺旋的双链

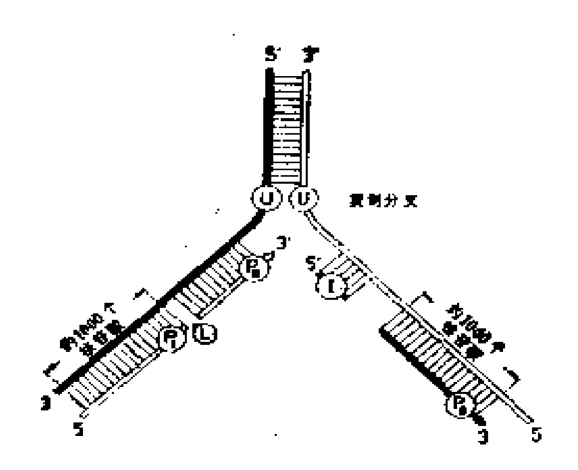


图 1-12 双链DNA 的半保守复制 这是一个极其复杂的过程，它包括了许多反应和控制步骤，图中示意了其中的最重要步骤。两条子链沿 5'→3' 方向上聚合起来。母代双螺旋的解旋通过一种解旋蛋白质(U)来实现。新片段的合成由 RNA引物引起，引物是在酶系统(I)的帮助下形成的，它后来又大概是在聚合酶I(P_I)的核酸酶功能作用下水解了。人们认为，链伸长到具有 1000 到 2000 个核苷酸的片段，主要是借助聚合酶 III 复合物(P_{III})来实现的。那些新生的所谓冈崎片段，必须由连接酶(L)连接在一起。在片段的3'端错误配对的核苷酸(且仅仅是这类核苷酸)，受到3'→5'核酸外切酶功能的作用，核酸外切酶很可能是聚合酶I复合物(P_I)，其活动主要与空隙填充和修补有关。诸如由 5'→3'核酸外切酶作用下的修补(通过它可以移去整个 DNA 片段)等其他特征，没有包括在这个系统中，因为它们对于新链的合成也许不那么重要。

中的每一条都被复制，从而导致两条(基本上一致的)螺旋，每一条螺旋中都包含一条母链。

(2) 复制从某个确定的生长点开始，可在该链的两个方向进行。借助于所谓的解旋蛋白质，使得双链解开。其中一些已被分离出来并经鉴别。它们上千倍地提高了解旋速率，从而带来比较快的复制分叉运动。同时，有必要把分子在某些部分解旋时所引起的扭力矩释放掉。这意味着，由核酸内切酶和连接酶来实现的链的断裂和修复机制中，会允许链段围绕着磷酸二酯键作居间旋转。

(3) 两条链的复制，用核苷酸残基沿5′→3′方向进行。*DNA*聚合酶只能以这种独特的取向方式利用单体，这不可能相伴地发生在两条链上。分辨力约为100Å的电子显微术已经揭示出，单链区域只存在于生长分叉的一侧，从而提示，只有取得了较大的伸展即满足5′→3′的进展之后，另一条链才合成完毕。

(4) 复制以不连续短脉冲的形式出现。在原核生物中，在一次脉冲中产生的片段长度约为1000到2000个核苷酸。它们是由生成引物而发动起来的，极小的RNA碎片充当了这种引物。以后，沿着这二条链进行着的*DNA*的复制片段，在复制分叉的后面由连接酶将其封接起来。

(5) DNA复制中所需要的各种功能，已由分离出各种特殊的酶并揭示它们的活性而得到了证实。特别是我们已鉴定了的几种聚合酶复合物（Ⅰ、Ⅱ和Ⅲ），它们包含了聚合功能以及几种链分解功能。在此特别有趣的是，聚合酶具有3′→5′核酸外切酸活性。它可以在生长链的3′端优先切除非碱基配对核苷酸。由于链生长仅仅出现在5′→3′方向上，所以这种核酸外切酶功能可以校对新合成链的片段。它的优化活性大约为聚合功能活性的2%。须把3′→5′核酸外切酶与5′→3′核酸外切酶区别开来，后者也是*DNA*聚合酶Ⅰ复合物的一部分，且很可能包含在切割修复中。它仅仅在5′端起作用，并在碱基配对区使得二酯键断裂，很可能使得5′端上脱落掉多至10个残基。因此，它可以除去一小段低聚核苷酸，而校对3′→5′的酶只在生长链的末端除去单个非碱基配对的核苷酸。

现在我们可以就RNA和DNA复制之间的重要差异作出结论，这种差异可以用两种机制的平均符号品质因子来表示。在*DNA*复制中，信息转移的精确性必须在连续的聚合作用中建立起来。然而*RNA*复制酶解决了这个问题，它显然取得了在0.9990

到0.9999之间的某个有限值$\bar{q}$。大致说来，任何连续的DNA聚合机制也应达到同样的精确度。

试管试验表明，突变体细菌噬菌体DNA聚合酶缺乏3′→5′核酸外切酶活性，这种酶还相对频繁地出现误差，大约每1000个核苷酸中出现1个。对于纯的*avian myoblastosis*病毒DNA聚合酶，已报道了类似的结果。不过，也发现了较低的误差率。例如，对比较小的、缺乏校对核酸外切酶活性的真核细胞*DNA*聚合酶所进行研究的结果表明，它所产生的误差率要比上述情况中低一个数量级(即0.1‰～0.2‰)。

(在原核细胞中)DNA聚合时出现的长为1000到2000个核苷酸的DNA片段，也可能直接与聚合酶功能的有限精度有关。显然，聚合酶不容易把由其自己产生的错误配对的末端伸展开，在没有核酸外切酶的情况下也已观察到这种现象。另一方面，如果存在3′→5′核酸外切酶，它将识别此种错误配对，并切除错误核苷酸。没有理由去假定：此阶段的最佳分辨能力大大不同于聚合中的最佳分辨能力。所以，校对又可以(最优地)把误差率减少另外3个数量级。另一方面，对误差的纠正不能拖延到任何后面的阶段，亦即不能在双链已经完成后才进行。尽管的确存在着运用5′→3′核酸酶活性的修复系统，但是它们不可能确定两链中是哪一链含有不正确配对的错误配对成员。更详细的动态校对机制已提出来了，并得到了实验检验。

因此，我们可以得出结论：*DNA*复制的最优平均符号品质达到了0.999999甚至更高，从而允许积累起相当于100万到1000万个核苷酸的信息(这取决于σ_m的大小)。这令人满意地注意到，这个数与已知的原核细胞中的基因组的大小(例如大肠杆菌，4×10^6个碱基对)吻合。而且，对于任何个体来说都完全不必达到这个极限。其他约束诸如在DNA噬菌体中的包装要求等等，可

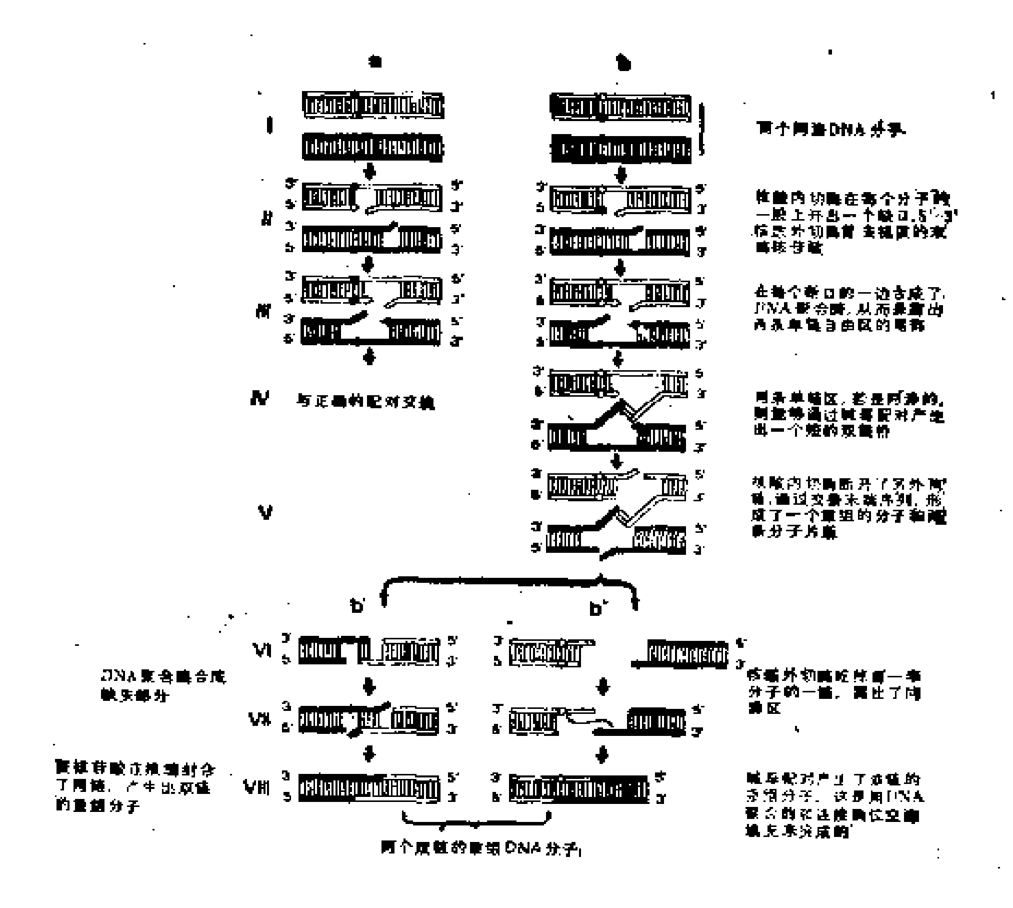


图 1-13　完整的双链 DNA 中可检出误差的遗传重组　原先提出这个模型是为了解释交换机制。它也可以适用于误差校正。符号 · 系指遗传上正确的核苷酸，0指有错误的核苷酸。因此，无论涉及到4种核苷酸中的哪一种，⁝ 总是代表了正确的互补配对，⁝ 代表了错误的(非互补的)配对．⁝ 也表示了互补但错误的核苷酸对。假定链的断口是错误配对所引起的（第II阶段)。那么，在一半情况中，3′—5′ 核酸外切酶的作用将校正误差：⁝ → ⁝（第III阶段)；而另外的一半情况中，是错误的核苷酸将自我互补：⁝ → ⁝。不过，不正确的(尽管是互补的)配对，并不是不变地处于某个简单的修复机制中。与正确拷贝的重组(第VI、VII阶段)将恢复原先的情形(第I和VIII阶段)，其中 4 个同源位置中，只有一个位置由不正确的核苷酸占据。所以，重复此程序可以导致平稳地减少误差，而不是使误差不可逆地固定在50%。这个系统指出，交换与完整的链中的检查误差相联系。此系统能够在 DNA 聚合酶的已知功能的基础上来加以理解，这并不排除其他假定的、甚至是更有效的机制。要完整地理解精确性问题，应当考虑到无性繁殖过程并要求有比现在更详尽的关于此种机制的知识。

能限制一个基因组的实际大小。正如对于RNA噬菌体那样，低于阈值的任何中等大小的基因组都是可能观察到的。

一个原核细胞的遗传信息量有一上限。正如任何超出10^4个核苷酸的单链信息容量的范围，需要具备新的双链模板和校对酶机制；由原核DNA复制机制确立的约10^7个核苷酸的新限度是不能被超过的，除非可以利用另一种进一步减少误差的机制。这种机制即遗传重组，它是自然界在原核水平上产生的。然而，使它完善起来从而把单个个体的遗传信息量提高到另一个范围，却耗去了20～30亿年之久。

被一切真核细胞所利用的基因重组过程，要求两个等位基因在其同源位置上被识别。由于酶促*DNA*复制的误差率对于每个核苷酸为低于10^{-6}，所以，未校正的错误是非常稀少的，并在两个等位基因的4个等价位置中不可能有一个以上的未经校正的错误。因此，对于重组物中的那些误差，哪怕原先是以完整双旋出现的，也还有更多的机会去认出并校正它们。图1-13描述了一种可能的系统。不过，对于重组机制，我们知之甚少，也不清楚在进一步减少误差率的过程中经历了多少步骤。事实上，正如对进化树分析所揭示的，误差率已经减少，这对于扩展遗传信息容量，并一直扩展到人的水平来说，它是一个重要的前提。

1.4.4 最初的复制单元

为了讨论生物信息的起源，我们必须从进化尺度的另一端出发，并对那些导致最初可复制的遗传结构进行分析。核苷酸所固有的物理性质影响辨别互补核苷酸与非互补核苷酸，其品质因子q不超过0.90至0.99。在对低聚核苷酸中的协同相互

作用的速度和平衡的研究基础上所进行的更详细的分析，已在其他地方有所报道。为了根据已知的自由能之差别来区别互补的和非互补的碱基配对，作为环境因子的、具有催化活性而又不受约束的丰富的蛋白质也许会有所帮助。不过，未受约束的蛋白质前体在一些情况下有利于互补相互作用，而在另一些情况下则有利于非互补相互作用。其中任何一种胜过另一种，都只能限于各种配对相互作用自由能的差别之中。互补配对相互作用的任何强化，就会要求那些有利于这种相互作用的特殊的酶有某种收敛进化。为了达到这一目的，那些酶自己必须成为自复制系统的一部分，这种自复制系统反过来又要求某种翻译机制的进化。

具有稳定信息量的最初自复制核酸结构——其最优 q 值在0.90到0.99之间——是类 *tRNA* 分子。然而，对于任何可复制的翻译系统，所需要的信息量至少要大一个数量级。正如我们对RNA噬菌体复制进行分析所了解到的，只有非常适应了的复制酶才能满足这种要求，但若无某种完善的翻译机制的帮助，就不可能进化到这种程度。我们今天所遇到的这些噬菌体，是以可利用这种机制为基础的进化的后期产物，要是没有这种机制，自然界就不会使一个核酸分子积累起那么多信息量。因此，在类tRNA结构层次上，核酸的进化有某种障碍，这种障碍类似于那些我们在进化的后期阶段发现的障碍，因此，为扩大信息容量就要有某种新机制。

于是，类 *tRNA* 或其前体似乎是“最古老的”复制单元，它们开始积累信息，并且被选作拟种即有相同基本结构的变异体。

第一个要求是水解的稳定性。已由类似于4.1中所描述的博奕模型表明，现在已知的tRNA的二级（和三级）结构（参见图1-14），是这种需求的直接进化结果。而且，这种结构的对称性

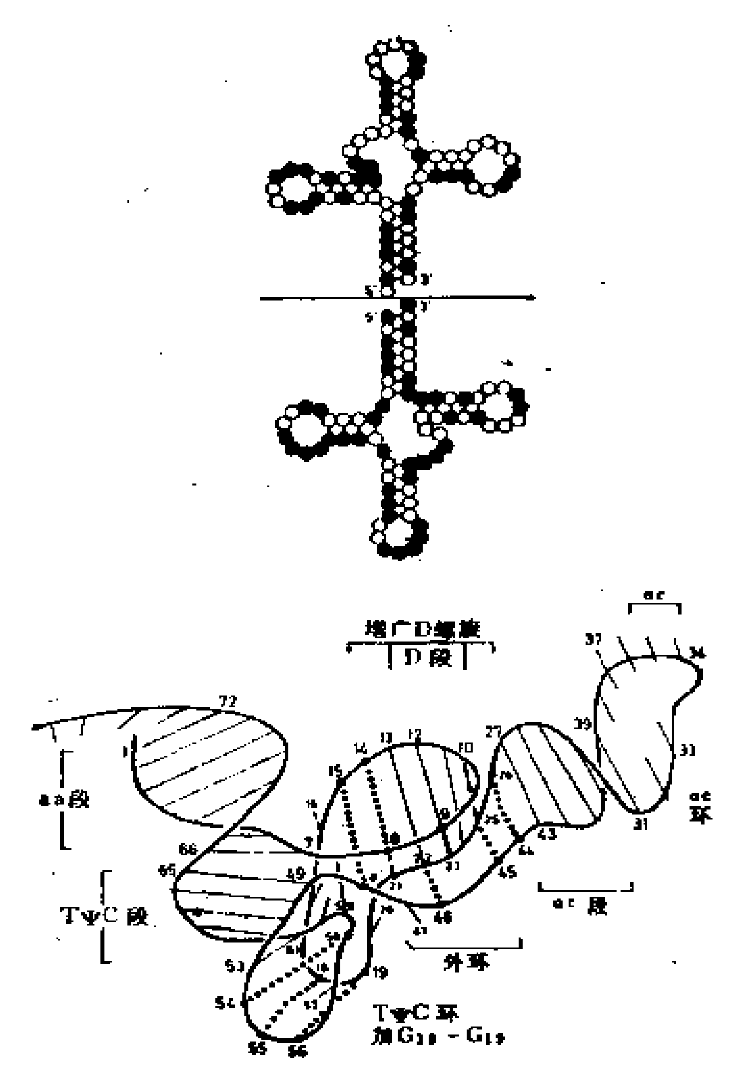


图1-14 功能RNA分子的对称性 这里以$tRNA^{Phe}$为例，这种对称性有助于专门适应酶的酶促单链复制。对称结构的正链和负链由普通的表现型特征来区别。尽管现存有机体中类*tRNA*是基因型编码的，但是它们的对称性也许仍然反映了单链RNA复制机制的原型，对此，正链和负链同样重要。这种对称结构在二级结构中最为明显(a)，但相应地也在三级结构中有所表现(b)。

反映了它的复制机制的最优化，按照(20)式，这意味着在此机制中正链和负链的行为是相当的，从而产生出最优性能。在*RNA*噬菌体的水平上，尤其是对于被选作在体外复制方面表现

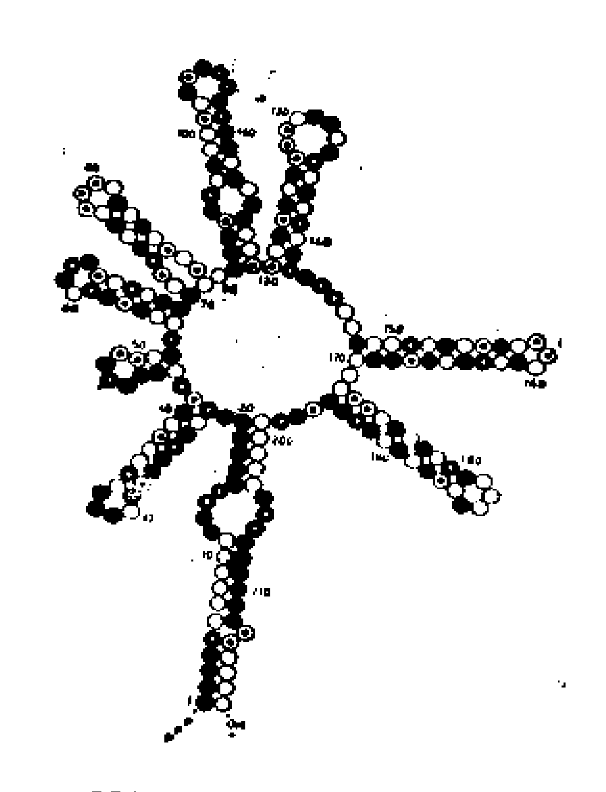


图 1-15 西格尔曼的QβRNA(正链)的中变体的"花瓣"模型 对称性要求不那么重要了，信息在此是由基因型规定的，基因型通过标准化的聚合机制加以复制。只是为了其表现型信息才选择噬菌体$Q\beta$的中变体，其中展示了一种被$Q\beta$复制酶所识别的最优目标结构。正链和负链都必定继承这种性质。在此"花瓣"模型中，结构对称性是十分明显的，尽管这种排列很可能并不代表自然的活性分子结构。按照图1-11所示的单链复制的力学条件，一个在合成中允许中间链段折叠的模型应该是具有优势的。

型上最有效的变异体，也有这种对称性，否则它们就不会携带遗传信息(图1-15)。

1.4.5 对超循环的需要

本文的目的就在于要表明，首先，分子进化中的突破必定是由几种自复制单元整合成协同系统所带来的；其次，能够进行这种整合的机制只能由超循环这类机制提供。这个结论又可

以基于以下论据从逻辑推理得出：

最初复制单元的信息量被限制在$\nu_{max}\lesssim 100$个核苷酸的范围内。为了构造翻译系统，要求若干个那种功能相似、专一性相异的单元。这种系统也许是从一种拟种中出现的，但是与之相当的伙伴应当同时进化。既不能把它们联接成一个更大的复制单元(因为存在误差阈)；也不可能由隔离作用来做到这一点，因为在分隔区间中的相当的自复制单元之间存在激烈的竞争。这种过程倒要求在所有自复制单元之中有功能联接，这可由如下品质来区别：

(1) 此种联接必须仍然允许每一自复制单元与其错误拷贝进行竞争，否则这些单元就不能保持其信息。

(2) 此种联接必须"切断"在那些自复制单元之间的竞争，那些单元应该整合成某种新的功能系统，并且在它们中间有协同作用。

(3) 于是整合的功能系统必须能够有利地与任何其他效率较低的系统或单元进行竞争。

只有自复制单元循环联接起来，才能够满足这3个要求，或者换言之，在自主的自复制单元本质之中的功能联接，必须具有自我增进的循环本性，否则它们的总信息量不可能可复制地加以保持。从而超循环组织对于具有较大信息量的整合自复制系统的聚集过程，看来是一个必要的前提，这正是翻译的起源所要求的。这种结论性的说法，即是后面对于已联接起来的系统进行更详细分析所要表明的东西。

如果我们问道："超循环的特殊之处是什么？"那么答案就是："它们是在组织的更高一级层次上的达尔文系统的类似物。"达尔文行为被看作是产生信息的基础。它的前提在于，自复制符号整合成自复制单元，自复制单元能够使自己稳定化以抵制

误差的积累。对于选择上稳定的自复制单元整合成更高级的组织形式，以复制出选择上稳定的行为，也有与上述相同的要求。只有此种循环联接——等价于在这个水平上加强自催化，才能满足上述要求。

表1-4总结了第4节的结果，它显示了达尔文系统中的一些基本的信息存贮阶段，而达尔文系统则被复制的各种存贮机制所促进。

对于讨论从单分子连续地进化成为整合的细胞系统的模型，这张表也将是有用的，正如第三部分所表明的那样。

表1-4 达尔文系统中的信息存贮的基本阶段

字符误差率 $1-\bar{q}_m$	优势 σ_m	最大字符量 ν_{max}	分子机制以及生物学例子
5×10^{-2}[a]	2 20 200	14 60 106	无酶RNA复制，tRNA前体，$\nu=80$
5×10^{-4}	2 20 200	1386 5991 10597	借助专一复制酶的单链RNA复制，噬菌体$Q\beta$，$\nu=4500$
1×10^{-6}	2 20 200	0.7×10^{6} 3.0×10^{6} 5.3×10^{6}	借助聚合酶酶促的DNA复制，包括由核酸外切酶校对大肠杆菌，$\nu=4\times10^{6}$
1×10^{-9}	2 20 200	0.7×10^{9} 3.0×10^{9} 5.3×10^{9}	在真核细胞中的DNA复制和重组，脊椎动物(人)，$\nu=3\times10^{9}$

a. 非催化的RNA复制从未满意地观察到，不过，表面催化或者由非专一适应的类蛋白(福克斯提出的)的催化，可以包括相应于表中所列的误差率的值。



抽象的超循环

拓扑方法被用来描述一类特殊的自复制反应网络：超循环。结果表明，超循环的性质足以稳定地整合那些包含在若干个自复制单元内的信息。对一些催化网络的研究结果表明，超循环组织是保持信息稳定性，并促使其继续进化的一个必要前提。本文中所运用的技术尽管为数学家们所熟悉，但是为了使得非数学家们也能接受这种逻辑论证，在论述时就要详细一些。

2.1 具体的问题

在超循环三部曲的第一部分中，已就分子水平上的达尔文系统得到了一些基本的结论，这可概括如下：

1．选择和进化的对象是拟种，拟种由（基因水平上）有密切关系的复制单元的某种分布构成，此分布集中在相应于最大选择价值的表现型拷贝（或一组简并拷贝）的周围。

2．以每复制单元的符号（核苷酸）数表示的这种主导拷贝的信息量，被限制在$\gamma_m < \frac{\ln\sigma_m}{1-q_m}$，式中$\sigma_m$（>1）是该主导拷贝

的优势，即对于分布的其余部分的平均选择优势，$\bar{q}_m$ 是平均符号复制品质。超过这个信息量阈值，将导致某种错误突变，即由于错误的逐渐积累而导致信息的溃散。

3．要达到数千核苷酸的稳定信息量，必须有高度进化了的酶复制机构。这样的信息量，对于编码诸如现存的RNA噬菌体中那样的几种蛋白质分子，是正好够用了。核酸中所固有的物理性质，允许可复制地积累起不多于50到100个核苷酸的信息。

基于以下论据，即认为诸如合适的催化表面或者甚至有类蛋白酶前体那样的环境因素可以大大改变那些数目，则可对上述 3 条陈述中的最后一条提出疑问。事实上，所给出的数字是从平衡态数据推导出来的，亦即从(协同的)互补核苷酸对于非互补核苷酸的相互作用自由能推导出来的。然而，我们仍然要考虑它们，要在自然界中实际上达到这种上限，只有存在合适的催化剂时或者通过退化处理才可能实现。由无酶模板诱导聚合实验得到的结果数字相当小。另一方面，环境的催化剂，除非它们自己成为选择上优化的系统之一部分，否则就不可能产生超过平衡态数据的符号辨识精确度。除了通过逐步的选择优化作用，决不会故意偏爱有功能优势的相互作用而超过无功能优势的相互作用。噬菌体基因组可能以单链*RNA*分子的形式进化，原因仅在于宿主细胞提供了相当先进的复制和翻译机构。它们是后细胞的进化产物，而不是前细胞的进化产物。正是它们的基因组所具有那么大的信息量是翻译起源时所需要的，这就是与一组酶有关的可复制的信息，这一组酶可以启动某个原始的翻译机构。因此从第一部分得到的最根本的结论是：开始翻译时要求把若干个自复制单元整合成某种协同系统，为建立翻译和复制机构提供充分的信息量。只有这种整合的机构，才能进一步增加精确度，从而允许相应增大信息量。

与归结到一个被ν_{max}的阈值关系所禁止的大的复制单元上去不同，人们是如何设想竞争性分子的某种整合作用呢？（注意，那些将被整合的单元必须保持对其突变体的竞争，以能继续进化，并不丢失其特定的信息。）让我们扼要地考察3种可能的选择：

1．共存。在没有稳定化相互作用时，只有属于相同拟种的个体，自复制单元才能稳定互容。拟种分布能够很好地为协同系统进化提供有利的起始条件。不过，它却不利于功能特征的进化。稳定拟种的耦合只由基因水平上的亲系关系来支配，这往往与功能需求不一致。所需要的是一组在表现型水平上互补的选择上等价的基因型。这种拟种分布不能满足这些选择标准。

2．隔离作用。把达尔文系统封闭在一个分隔间中，也无助于解决这个问题。空间隔离的主要结果是加强了竞争，原因在于限制了生活空间和代谢供给。因此，一个分隔间将只能进一步稳定某个给定的有选择优势的拟种；它将不利于其他伙伴的进化，按照功能标准，那些伙伴是与此拟种相当的，这种进化要求协同单元在基因水平上有趋异作用。然而，对于一个通过功能连接而确立了稳定协同作用的系统（参见第三部分），一个分隔间却可以对它提供优势。当然，象今天活细胞那样的更为高级的分隔间，仅由每个复制亚单元的一个（或几个）拷贝与复制整个分隔间的机构一起构成，它们要求符号品质q_m，符合ν_{max}关系这个q_m是适合于总信息量的。换言之，它们与一个完全连接起来的单元服从相同的限度。

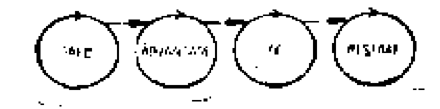
3．功能连接。它们或是提供了复制的相互催化增强作用，或是提供了某种结构稳定化作用，从而影响功能上协同的那些伙伴的选择。对这样的连接进行更为深入的考察，是本文的主

题。

让我们进行在第一部分中引入的计算机博奕，以助我们的直觉认识。博奕第一部分的目的在于揭示出需要使符号复制品质适合于被复制语句的信息量。在第2部分我们假定：平均品质因子 $\bar{q}_m$ 不足以使整条语句以复制单元形式进行稳定复制，但足以使得象单词那样小的复制单元进行稳定复制。它与早期进化的自然状况有关，在此核苷酸应有的物理力也许已经充分满足稳定的类$tRNA$分子(＝单词)的进化，但不允许构造起哪怕是原始的翻译器(＝整条“有意义的”语句)。相应地，所编的计算机程序正好可复制出这样的单词；它们使用的误差率足以抵制误差积累而保证其稳定性。把原先的博奕改变一下，让我试着去建立个词的一种简单共存。为达此目的，我们赋予语句中所有正确的词以相同的选择价值，而任何词中的一个错误相对于正确的词来说是不利的，它们相差一个给定因子(每比特)跟以前一样，词可以进行复制，其总数限制为N个拷贝。这种新的博奕与原先的博奕不同之处在于，单个词现在是作为独立的复制单元在起作用。表1-5显示了一些典型的结果，尽管所有的词具有相同的选择价值并能处于优势地位与其错误拷贝竞争，但总体上语句是不稳定的。4个词中只有1个能赢得竞争，但是无法预见这4个词中哪一个词实际上将赢得竞争。人们可以用同义反复，即“生存者生存”来描写这种情形。“适者”一词不是别的，只是竞争仅有的结果。

表1-5 选择上等价的复制单元之间竞争的博奕

博奕的目的在于保持语句的信息：



每个词就是一个符号表示的复制单元。所有的词都具有完全相等的选择价值。每比特的选择优势为 2.7。每个字母由5个二元字符构成。

	字符突变概率$(1-\bar{q})$ [%]	字符数 ν	词品质因子 $Q=q^2$	误差期望值 E
TAKE	3.15	20	0.53	0.63
ADVANTAGE	1.4	45	0.53	0.63
OF	6.3	10	0.53	0.63
MISTAKE	1.8	35	0.53	0.63

因为在词之间不存在耦合，所以，选择了一个词的时候结束博奕。所有词的选择价值是简并的，所以4个词的生存机会都是相等的。由于平均误差概率高（约每比特2%），所以此语句作为一个整体(125比特)不是一个稳定的复制单元。

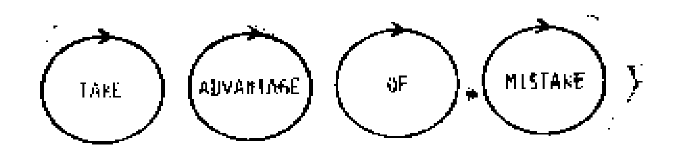
下面列出典型的10次博奕结果。"X" 标志被选择的词，而其余的词都消亡了。同时还给出了完成选择后的世代数。

博奕	TAKE	ADVANTAGE	OF	MISTAKE	世代
1	×				12
2			×		15
3		×			19
4		×			23
5	×				10
6				×	20
7			×		9
8		×			13
9			×		22
10		×			26

被选择的词 ADVANTAGE 的误差分布。黑线条类似于泊松分布 $\frac{\varepsilon^k \cdot e^{-\varepsilon}}{k}$；式中 $\varepsilon=\nu(1-\bar{q})$是在该词（$\nu=45$比特）中的某个误差的期望值(误差涉及到单个数位。错误字母与正确字母的差异只在于它们的5个数位中有一个不同)。

无误差	单误差	双误差	三误差
ADVANTAGE			
ANVANTAGE			
ADVANTAGE			
ADVANTAGE	ADVANTAG!		
ADVANTAGE	ADVANDAGE		
ADVANTAGE	ADVARTAGE		
ADVANTAGE	ADVINTAGE		
ADVANTAGE	ADFANTAGE		
ADVANTAGE	ADZANTAGE		
ADVANTAGE	AHVANTAGE		
ADVANTAGE	AHVANTAGE	ADVANDACE	
ADVANTAGE	AHVANTAGE	ADVANDAIE	
ADVANTAGE	AFVANTAGE	ADVARXAGE	
ADVANTAGE	AFVANTAGE	ADVINTBGE	ITVANTAGU

将此博奕再修改一下，在有关系的词之间引入某种功能连接，也就是说，只要形成了有意义的序列，则一个特定的词就对下一个词的复制提供催化帮助：



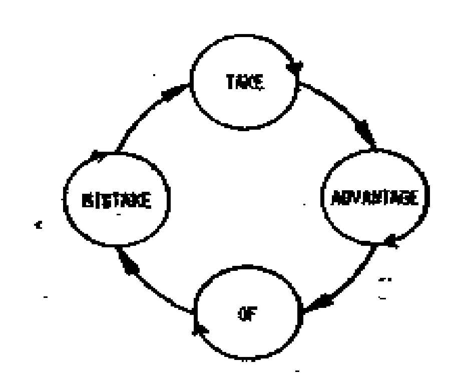
此耦合正比于催化剂(即在计算机存贮中的特定词的样本)的群体数。换言之，复制按照速率定律得到促进：

$k_i x_i$　　以及

$k_i x_i + k_i' x_i x_{i-1}$　　当 $i=2,3,4$

x_i 和x_{i-1} 分别是群体数，在本例中是指存贮在计算机存贮器中的词。这种博奕变异的结果，通常是固定语句中最后一个词"MISTAKE"，而其余的词都消亡了。只有在耦合相对弱，所选择的特定 k_i 值充分大时，我们发现，相应的词 (i) 的确可以从中生长起来，这代表了(本质上)独立的竞争者之间的选择。结果是此序列中最后一个词得益于所有的耦合（只要耦合项处于主导地位），此结果也许是令人惊讶的。人们或许会期望，整条语句至少有一个稳定范围。对于某个大小的群体数，这肯定是对的。如果在词的位置方面速率参数值遵从某个次序的话。然而，第7节的定点分析将表明，即使在那些特殊条件下，也只有链中的最后一个成员将与总群体数成比例地生长，所有的其余成员基本上采取恒定的群体数，与总群体大小无关。因此，在生长群体内，最后一个成员的相对丰度会发生剧烈的变化，直到系统再度达到某个范围，此时只有这个丰度大的成员保持稳定。分子进化过程中，个体的群体数通常表现出那样的剧烈变化，亦就是从一个突变体变到能被检测的(大于)数十亿个拷贝那么大。所以，在我们的博奕中所获的结果，原来也许正好代表了自然界中实际会发生的情况。

线性耦合(如果它始终是起作用的)把所有的优势都传给序列中最后一个成员，这一事实强烈地暗示了对这个问题的一个



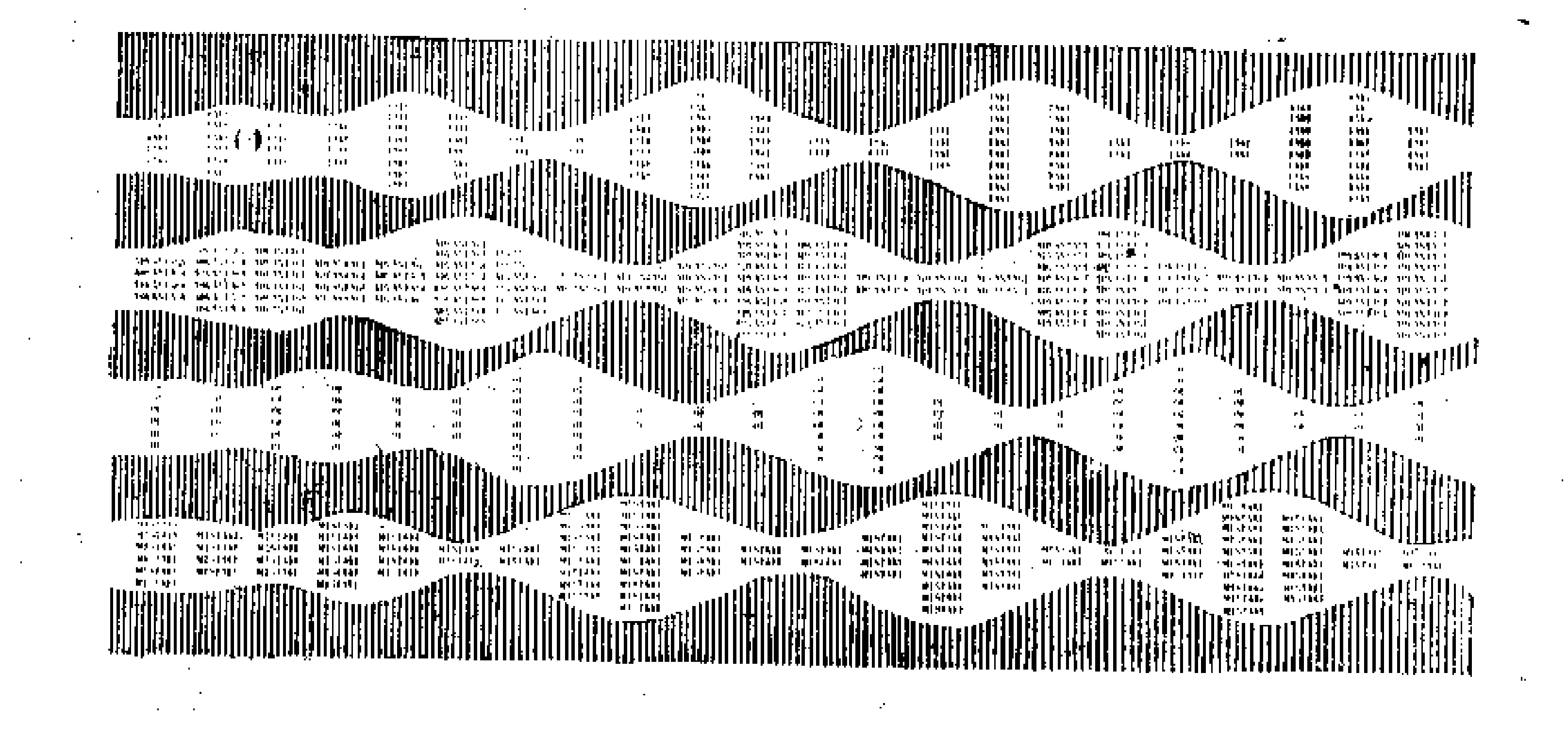


图1-16 语句 TAKE ADVANTAGE OF MISTAKE 中的每个词代表一个自复制单元。语句的信息由词中的超循环耦合来稳定。图中的每个词代表了存贮在计算机中的10个拷贝。在同一列的词处于相同的时间，其间的间隔沿水平轴不连续地变化。从起始均匀分布产生出来的振荡，保持了某个相从一个词转移到另一个词。4个词中每个词的误差分布都是稳定的，并且都与表 1-5 中所示的种类相同

可能解，即耦合应该形成一个封闭的环：
因此，耦合所致的增强作用将通过此序列中所有的词而循环地涨落。这条语句，实际上就是自动地得到这种通过“mistake”一词的循环重叠。由于一个词是一个催化循环（即一个自复制单元），所以据我们在第一部分中所引入的定义，此系统代表了一个二级超循环。图1-16代表了此博奕的结果。所有的4个词显示出群体数周期变化的一种稳定的稳态表现。不同词的选择价值不必相同，这对任何现实的系统来说看起来是极不可能的。而且，每个词都由一个突变体的稳定分布来代表。除非其中某个词被涨落突变消灭了（在有充分大的拷贝数时这是极不可能的），否则群体数将连续地变化。换句话说，整条语句的信息是稳定的。

2.2 动力学系统的一般分类

2.2.1 定义

以下各节中，我们将对动力学系统尤其是那些在前细胞自组织中具有重要性的动力学系统，进行更严格的数学分析。为确定哪些系统是相关的，我们就必须考察不同类的反应网络，其中包括非循环反应网络和循环反应网络。进化过程可以由微分方程组唯象地描述，这在第一部分中已有一个特例。于是，名词动力学系统与一给定微分方程组解的曲线的完整流形有关。

让我们考虑一个一般的动力学系统，它由n个一阶自控常微分方程来描述：

$$\frac{dx_i}{dt}=\dot{x}_i=A_i(x_1,\ldots,x_n,\ k_1,\ \ldots,k_m;\ B) \tag{30}$$

$$i=1, 2, \ldots, n$$

稍后我们将把我们的分析推广到一些非自控系统，对于它们$\Delta_i=\Delta_i(t)$。

跟以前一样，x_i代表群体变量，它们通常将涉及到自复制的大分子集合体。常量$k_i(i=1, 2, \ldots, m)$是参数，构成这些常数的可以是：基本过程的速率常数，可逆并且快速建立的反应步骤的平衡常数，以及那些作为合成大分子(高能)原料的分子浓度(假定这些浓度是缓冲的，因此可以作为与时间无关的值包括进来)。x值和k值的集合，可以分别表示为浓度空间或参数空间中的列向量：

$$x=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad 以及 \quad k=\begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{pmatrix}$$

我们用B标示给定解的曲线集合的起始条件，本例中就是起始浓度集合x_0。

按照第一部分中所运用的程序，我们把函数Δ_i分解成3项：

$$\Delta_i=A_i-\Delta_i-\phi_i \tag{31}$$

A_iS 构成了对化学速率的所有正贡献，即表示了x_i变量的"放大作用"，而$\Delta_i s$包括了所有的负速率项，即类似于大分子物种的"分解"。ϕ_i则代表了某种流，它可以影响组分i的稀释或缓冲作用，这取决于施于系统的外部约束。差$A_i-\Delta_i$可以称作净生长函数Γ_i。参照达尔文系统(参见第一部分)，特别地，Γ_i可由$W_{ii}x_i+\sum w_{ik}x_k$给出，如果对所有物种从$k=1$到n求和，它就类似于超额生长函数$E=\sum_{k=1}^{n}E_k x_k$。

2.2.2 无限制生长

去掉选择约束便导致新的微分方程组

$$\dot{x}_i = \Gamma_i(x, k, B) \tag{32}$$

它描述了下面所谓的"无限制生长"情形。这个术语对于作为一个整体的系统是有代表性的；对个体成员来说，它还可以包括衰退的或平稳的行为。

假定我们可把 Γ_i 表示为各种浓度 x_i 的多项式（作为一种近似还可以适用于无理式或者多项式的比），那么，通常可能找到 Γ_i 的首项，它们在一定浓度范围内处于主导地位。这些首项通常是x_i的一个给定幂的简单多项式。从而它们决定了该系统的动力学行为。

图1–17例示了简单情形：$\dot{x}=kx^p$。此范例的解已归一化到$x(0)=1$和 $\dot{x}(0)=1$。正如图的说明所勾划的，整个解曲线族可以划分为 3 类，它们被限制在浓度–时间图中的不同 区 域。让我们考虑3个典型例子，它们对我们后面进一步的讨论将具有特殊意义(参见表1–6)。

(1) 解曲线1代表了一个具恒定(正)生长速率的系统。群体变量 $x(t)$ 随时间线性地增加。此条解曲线是图 1–17 的区域B中曲线族的一个例子。这些曲线在无限时间处无限生长。具有缓冲的总反应物浓度的不可逆生成反应，可以当作最为普通的例子。在生态小生境中以独立资源为生的自复制物种，可以调节它们的生长速率为恒流的或恒定的食物产生率，从而构成了与群体大小无关的生长行为的另一个例子。

(2) 解曲线2是线性群体变量的生长速率的结果，它显示了x随t指数增加，正如在第一部分中表明的，这是典型的达尔文行

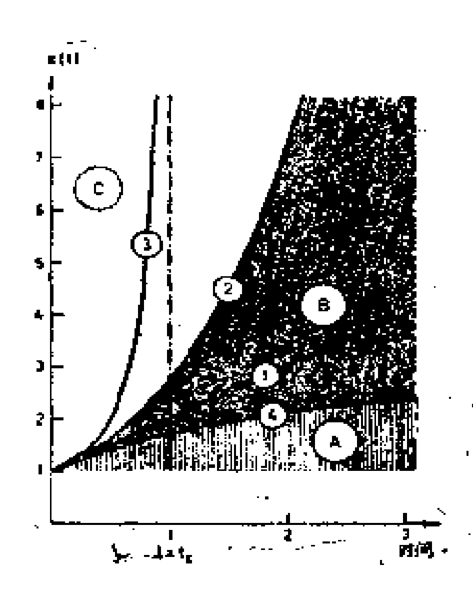


图 1-17 不同类型的生长可以与单项的生长函数 $\Gamma(x)=dx/dt$(归一化到，对 $t=0$，有 $\Gamma=1$ 和 $x=1$)相联系。区域 A 不包括任何可由简单的单项式 $\Gamma=x^p$ 表示的生长函数。在这个区域，总群体数 $x(t)$ 在时间 t 趋于无限时仍保持有限。在区域A与 B 之间的边界线是由生长函数（曲线 4）$\Gamma(x)=e^{1-x}$ 给出的。区域 B 包括了从 $-\infty<p<1$ 的所有单项式 $\Gamma(x)=x^p$。在这个区域中的曲线 1 作为等价于线性生长的恒定生长速率($p=0$)的例子。指数生长($p=1$)给出了在区域 B 和 C 之间的边界线(曲线2)。在区域 B 中的群体数，只有在无限时间处才达到无限。在区域C 中，在无限时间处它们显示出一些奇点。作为一个例子，曲线 3 显示出双曲线生长($p=2$，奇点在 $t=t_c=1$ 处)。

为。曲线2还成为区域 B 和 C 的分界线，也就是在无限 t 时达到无限和在有限 t 时达到无限的函数的分界线。

(3) 解曲线3是在有限时间处 $[t_c=(kx_0)^{-1}]$ 具有奇点的函数的例子。在这个例子中，生长速率被假定为与群体变量的平方成比例。整个区域C 的特征可以称为“双曲线生长”。当然，在任何现实的、有限的世界中，一个群体决不可能无限生长，因为可用资源是有限的，从而总有约束条件制约生长限度。那些引起假定存在着奇点的现象，仍将导致与达尔文系统大不一样的

行为。

有鉴于此，我们可以更一般地定义生长函数的"生长度"p，它对于分类将是有用的。跟以前一样，p_i 是在生长函数 Γ_i 中的首项的幂。因此，一个 n 维的动力学系统可以用一组 p_i 值$(p_1, p_2, \dots, p_n)$来表示。当我们有幂 p_i 的均匀分布时，即

$$p_1 = p_2 = \dots = p_n = p \tag{33}$$

我们称此系统是"纯净的"。否则我们就将面临"混合的"系统，它们可以据其p_i值分布来分类。显然，与"混合的"系统相比，对"纯净的"系统的分析要容易得多。

2.2.3 约束生长和选择

在现实中，我们总可以遇到对生长有一定限制的约束条件。对于实验研究，我们必须为复制的条件提供保证。因此，有必要把这些条件形式化，并把它们包括在理论处理中。

在不可逆热力学中，我们倾向于选用有利于热力学描述的选择约束，例如恒定的广义力或恒定的广义流。在这里的分析中，我们必须调节这些约束，使之适合于选择和进化所要求的条件，它们在自然界中能被具体化。在(31)式中运用的约束条件ϕ_i，对于任何直接的分析来说，都太一般化了。通常说来，我们可以把专一的与非专一的约束加以区别。在前一种情形中，约束专一地对一种物种或几种物种起作用，而在后一种情形下涉及到调整某个总流 ϕ。于是，在所有群体变量中的变化都与它们的实际值x_i成比例：

$$\phi_i = \frac{x_i}{c}\phi \tag{34}$$

在实践中运用某个连续稀释流可以把非专一的选择约束引

入动力学系统中。由此总浓度 $c=\sum x_i$ 是可以加以控制的。关于 c 的相应的微分方程：

$$\dot{c}=\sum_{j=1}^{n}\dot{x}_j=\sum_{j=1}^{n}\Gamma_j(x)-\phi \tag{35}$$

它满足稳定性条件：$\dot{c}=0$，此时这个流已调节到抵消了净的超额产生，即

$$\phi=\phi_0=\sum_{j=1}^{n}\Gamma_j(x) \tag{36}$$

这种被称作"恒组织"的选择约束，在前面就已引入了，并在第一部分中得到了运用。在后面的各节中，将反复用到条件(36)，以有利于对选择过程进行一般分析。我们还考察了一些其他约束。正如在下一节将见到的，选择过程和进化过程的重要特征对于约束条件是相当不敏感的。(这些当然总是以定量结果来反映的。)

恒组织条件导致下列动力学系统的微分方程：

$$\dot{x}_i=\Gamma_i(x)-\frac{x_i}{c_0}\sum_{j=1}^{n}\Gamma_j(x),\quad i=1,2,\ldots,n \tag{37}$$

式中 c_0 表示总浓度的稳定值，这可以由把流的大小调节到 ϕ_0 得到保持。

与图1-17的讨论相联系的 $p=0$，1和2这3个简单的生长函数，其个体选择行为详列在表1-6中：

(1) 恒生长速率，它相当于群体随时间线性增长，在恒组织约束下使得系统中所有的伙伴稳定共存。优势突变体的生长改变了这个稳定比率，但不导致整个系统变得不稳定。

(2) 线性生长速率，对应于群体按指数增长，结果导致竞争和最适者的选择。有利突变体的出现动摇了已建立的群体，并取而代之。

表1-6 动力学系统 $\dot{x}=\Gamma-\phi$ 中，在恒总组织的选择约束下的生长速率和选择行为

p	无限制生长			在恒组织约束下的长期行为	
	生长速率 $\Gamma(x)$	解曲线	生长类型	$\bar{x}$	选择行为的类型
1	k	$x=x_0+kt$	线性	$\bar{x}_i=k_ic_0/\Sigma k_j$	物种共存，无选择
2	kx	$x=x_0exp(kt)$	指数	$\bar{x}_k=c_0$ $\bar{x}_i=0$ $k_k-k_i>0$ $i\neq k$	竞争，导致总观的"最适者"物种的选择
3	kx^2	$x=x_0(1-kx_0t)^{-1}$	双曲线	$\bar{x}_k=c_0$ $\bar{x}_i=0$ $k=1,2,\ldots n$ $i\neq k$	竞争，旨在处于区域最优化"一旦则永存"结果

（3）非线性生长速率（$p>1$），特征是双曲线生长，这也导致选择，而且比（2）中提到的达尔文系统的选择更激烈。不过，具优势速率参数的突变体，一般不太可能生长起来并动摇一个已建立的群体，因为其选择价值是群体数的函数（例如，对 $p=2$，有 $W=x$）。任何已建立的具有有限x的群体优势是如此之大，以致任何单突变体拷贝都难以与之挑战。选择则代表了一种"一旦则永存"的结局。若干种物种的共存，在此需要形式非常特殊的协同耦合。

上述例子颇具代表性。我们可以根据其选择行为，把系统划分为共存的或竞争的系统。在一定的系统中，我们会遇到一种以上的行为。

2.2.4 生长系统中的内平衡

恒组织约束大大简化了对动力学系统的分析，但另一方面又使得它局限在净生长为零的系统。在本节中，我们将致力于

扩展其适用范围。主要问题在于要找到，根据对相应的稳态分析所获的结果，在什么条件下以什么方式能够对生长的系统作出预见。为做到这一点，我们引入非专一的、依赖时间的选择约束[(34)式]：

$$\dot{x}=\Gamma_i(x)-\frac{x_i}{c(t)}\phi(t) \tag{38}$$

$c(t)$ 或 $\phi(t)$ 都可以自由地加以选择。不过，另一个函数则分别由如下的微分方程或积分方程来决定：

$$\phi(t)=\sum_{i=1}^{n}\Gamma_i(x)-\frac{dc}{dt} \tag{39}$$

或

$$c(t)=c_0+\int_0^t\left\{\sum_{i=1}^{n}\Gamma_i(x)-\phi(\tau)\right\}d\tau \tag{40}$$

现在引入标准化的群体变量 $\zeta=\frac{1}{c}x$ 是合适的。于是得到如下形式的微分方程：

$$\dot{\zeta}_i=\frac{1}{c(t)}\left\{\Gamma_i(x)-\zeta_i\Sigma_j\Gamma_j(x)\right\} \tag{41}$$

我们立刻可见到，$\dot{\zeta}$ 并不明显地依赖选择约束 $\phi(t)$。不过，倒是通过 $c(t)$ 有一种隐含的依存关系。因此，我们考虑一些明显的例子，以把我们的一般分析深入一步：假定净生长函数 $\Gamma_i(x)$ 是 x 的 λ 次齐次函数。虽然这个条件看起来是非常有限制的，但是我们将看到，几乎所有重要的模型系统都将与这种函数对应，至少在一定的边界条件之下是这样。x 的齐次性导致了与在无限制生长的系统（参见1.5）中所要求的一定生长度p（$\lambda=p$）相同的条件。现在，变量的变换是相当普通的：

$$\Gamma_i(x)=\Gamma i(c\xi)=c^{\lambda}\Gamma_i(\xi) \tag{42}$$

从而速率方程为

$$\dot{\xi}_i=c^{\lambda-1}\left\{\Gamma_i\ (\xi)-\xi_i\sum_{j=1}^{n}\Gamma_j\ (\xi)\right\} \tag{43}$$

对此方程进行简单的考察，可获得两个重要的结论：

如果 $\lambda=p=1$，即第一部分中所讨论过的达尔文系统，那么对 c 的相关性就不存在了，并且就相对群体变量 ξ_i 而论，在生长系统和稳定系统中，不仅长期的行为而且解曲线都是一致的。

如果 $\lambda=p\neq1$，假若 $c(t)$ 既不为零也不为无限，那么关于 ξ 的长期行为与恒组织的稳定系统是一致的。从而对于全部现实的具有齐次的净生长函数 Γ_i 的系统，正如下一节中获得的，对ξ 空间进行固定点分析的结果，将对于生长的群体仍然有效。

在固定点分析的那一节将表明，把上述的后一种结果推广到其他种类的生长函数是可能的。内平衡大大简化了对复杂的动力学系统的分析。在许多情况下，这些结果与那些稳定条件下的结果是一致的或相似的。如果我们对系统的选择行为进行分析，那么上述结果也就是实际上遇到的条件。在下一节，我们将更深入地考察在这些条件下的各种动力学系统。

2.3 自组织反应网络的固定点分析

2.3.1 合适的分析方法

分析各种分子自组织过程时，我们自然更关心最后的选择结果，而不是着眼于动力学过程的详细求解。因此，在这一节中，我们并不需要由满足某个微分方程组的整族解曲线提供的

全部信息。因此，我们选用的方法是固定点分析，因为它最适宜于对选择行为进行比较分析。仅仅在某些情况下，我们才考虑更为高级的技术，例如考察完全向量场。

现在，对于研究动力学系统的长期行为，固定点分析是一种常规的技术，它可以在数学教科书或论文中找到。固定点分析也已被运用于经济学、生态学模型以及远离平衡的化学反应等方面的问题。最近有一份进展报告概述了它在现阶段的发展。

2.3.2 拓扑特征

让我们借用一张山区地图来形象地谈这个问题（参见图1-18）。二维地图的等高线使我们对三维的图景有一个大致的印象。固定点分析正是用来讨论这种问题的。地形相当于一个势能面，动力学系统沿这势能面运动。在大多数情况下并不需要这种势能面的完整知识，因此，当我们在一未知区域要确定自己的方位时，运用“固定点图”比起运用测量图，则要简单得多。一般说来，“固定点图”专门标明了当地最高地点和最低地点的位置，例如山峰、关口和山谷等。这些位置被称作源头(发点)、山脊（鞍点）和渗坑（收点)。这些特殊的点就是势能场的不动点。它通常还需要把一个山谷与另一个山谷隔开的分水岭线包括进去(图1-18)，这些通常称作“分界线”。一幅包含分界线的固定点图，足以预见在图上从一定点出发的轨道将引向何处。轨道是下降得最陡峭的路线，在地形中这种路线是水流要采取的。另一方面，地球表面的重力势场没有我们在自组织动力学系统中所遇到的场那么复杂。因为地球上的水流总是要趋向诸如湖泊这样的收点，而自组织的动力学系统则可以表现出更复杂的行为。例如，存在一些称作为极限环的情形，以我们举例时所

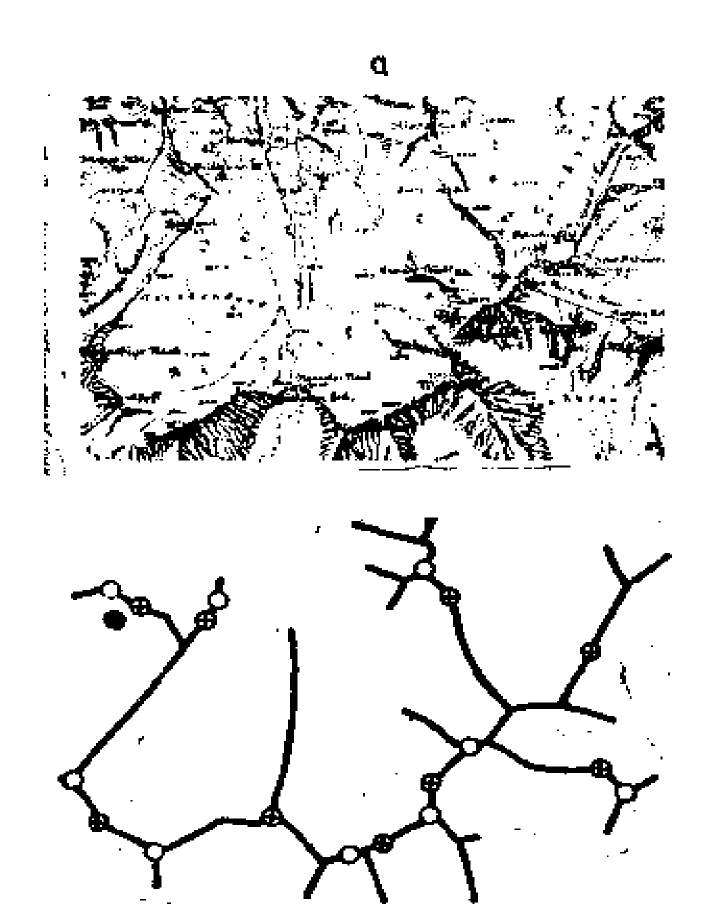


图 1-18 (*a*) 拓扑图是对地形的一种抽象表示 点之间的连线是等高的。图中是阿尔卑斯的一个地区。
(b) 固定点图是对拓扑图的进一步抽象 不固定点(⊖)表示水源或山峰(发点),⊕表示山脊点(鞍点),●表示渗坑(收点);图中黑线标志分界线。

用的话来说，在此处水流不会在某一点停顿下来，而将沿着一条由势场形状所确定的封闭线不停地循环。还有更为奇特的情形，数学家实际上称其为“奇异吸引子”，它们表示的是诸如非周期轨道之类的东西。比起收点来，吸引子是一种更为一般的表达方式。它不仅包括了收点，而且包括了稳定封闭的、非周期的轨道。

在一幅固定点图中，所考虑的整个地区可以划分成一些区域，通常称作洼域，它们属于单个的吸引子。这些区域的边界是分界线。所以，在洼域内的所有点，水都流向相同的吸引子，吸引子当然处在该洼域之内。

现在，让我们更为精确地描述用数学术语进一步讨论时所必要的量和表达式。动力学系统的固定点或不变点被定义为这样的点：在此所有的浓度或群体变量 x_i 是不随时间变化的。所以，一阶时间导数为零：

$$\dot{x}_i=0 \qquad i=1,\ 2,\dots,\ n \tag{44}$$

由此决定了属于一给定动力学系统的所有固定点的位置。当群体变量中的所有随机涨落都被严格地抑制了的时候，从固定点出发的动力学系统，其整合作用很普通地导致与时间无关的恒定群体。系统在一给定固定点上对小的浓度变化的响应，为这些点的分类提供了很好的工具。它可以由一组具有倒易时间常数 ω_k 的简正模来描述，ω_k 是某个线性微分方程组的本征值，此线性系统是参考点附近的非线性系统的最合适代表（参见2.3.4）。相应地，我们可以区分4类主要的固定点。

（1）稳定固定点或收点，即当地的最低点。所有本征值 ω_k 实部为负，因此，沿浓度空间的所有可能方向上的涨落，由内部的反作用力来补偿。在化学中，收点相应于封闭热力学系统中的化学平衡态和开放热力学系统中的稳定的稳定态。

（2）鞍点，至少有一个不稳定方向。在此至少有一个 ω_k 值具有正的实部。因而在这个方向上的小的微扰或涨落都会引起某种倾向于增大此涨落的力。结果是动力学系统将离开此鞍点。

（3）发点，代表了当地的最高点。它与鞍点的不同仅在于，它在一切方向上都是不稳定的。所有的 ω_k 值都有正的实部。

（4）另一类不动点，它们不可能以线性方式来彻底分析。其

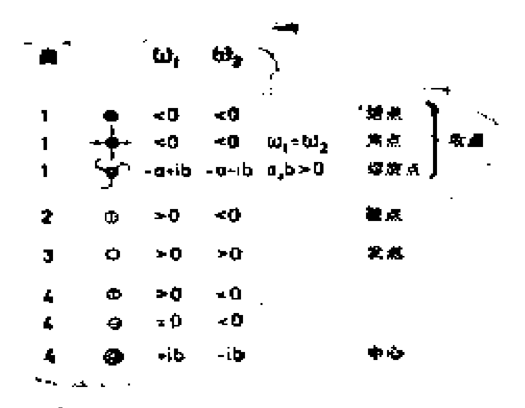


图 1-19 运用符号来区分各类不同的固定点：第1类是稳定的固定点或收点；第2类是鞍点；第3类是发点；第4类是不稳定的固定点，其中包括具有实部为零的本征值。这些例子涉及到2维动力学系统。

中一些频率 ω_k 显示出实部为零，而非线性贡献又可以改变它们的本性。中心点是一个例子，它们是由纯虚数的本征值来定义的，其轨迹表现同心轨道流形。在本文中将碰到这种情形。

在"充分长"的时间以后（充分长的时间意味着一个周期比起动力学系统的最大时间常量要大得多），每一个现实的动力学系统（即一个没有涨落的外部抑制系统）将趋近某个吸引子。所以，选择的结果总是要与浓度空间中的某个吸引子吻合。

选择过程的最终结果，即对应于某个稳定的定态，或是对应于某个连续地周期性地变化的状态族。在一些非常罕见的情形中，也可以在一定的状态集合中出现非周期变化。微分拓扑学中，对所有这些稳定的或亚稳的终态情形，有一个共同的一般表述，即称之为动力学系统的"吸引子"，其中包括稳定点、封闭轨道和非周期线。在一定洼域中，选择过程的结果是趋向一个吸引子，它与特定的起始条件无关。

2.3.3 合适的空间:浓度单形

浓度变量或群体数生成 n 维开放空间 R^n: $\{x_1, x_2, \ldots, x_n; -\infty < x_i < \infty, i=1, 2, \ldots, n\}$,其中只有一部分是物理上有意义的:

$$X^n \subset R^n; X^n: \{x_1, x_2, \ldots, x_n; x_i \geqslant 0, i=1, 2, \ldots, n\} \tag{45}$$

对所有的浓度变量求和,得出一个非负的有限总浓度 c:

$$c = \sum_{i=1}^{n} x_i, \quad 0 \leqslant c < \infty$$

它被用来进行归一化:

$$\zeta_i = \frac{x_i}{c}; \quad 0 \leqslant \zeta \leqslant 1, \sum_{i=1}^{n} \zeta_i = 1 \tag{46}$$

根据这些变量的性质,对于任何给定的 $c=c_0$, X^n 可以被同构地映射到一个单位单形 S_n 上(相应的空间记作 S^n):

$$c=c_0: X^n \Longleftrightarrow S^n: \{\zeta_1, \zeta_2, \cdots, \zeta_n\} \tag{47}$$

单位单形 S_n 是一个具有 n 个隅角的正多边形,它相应于由 $\sum_{i=1}^{n} \zeta_i = 1$ 所定义的 $(n-1)$ 维子空间。它有单位边长,边代表变量 ζ_i 的坐标轴。作为一个实例,图 1-20 中显示了一个 S_3。S_3 图为化学家所熟悉,它代表三元系统。

作为 (46) 式的结果,与 X^n 比较,在单位单形上的动力学系统已经丢失了一个自由度。换言之:由于归一化,变量 ζ_i 总是与一个定值 $c=c_0$ 有关,从而在变量中引入了一种线性关系。

最后,我们将要强调一下在 X^n 图和 S^n 图之间的差异,当我们把对于不同 c_0 值所获得的结果进行比较时,差异就变得明显了。由于归一化,单形 S_n 的大小是固定的,而物理上可

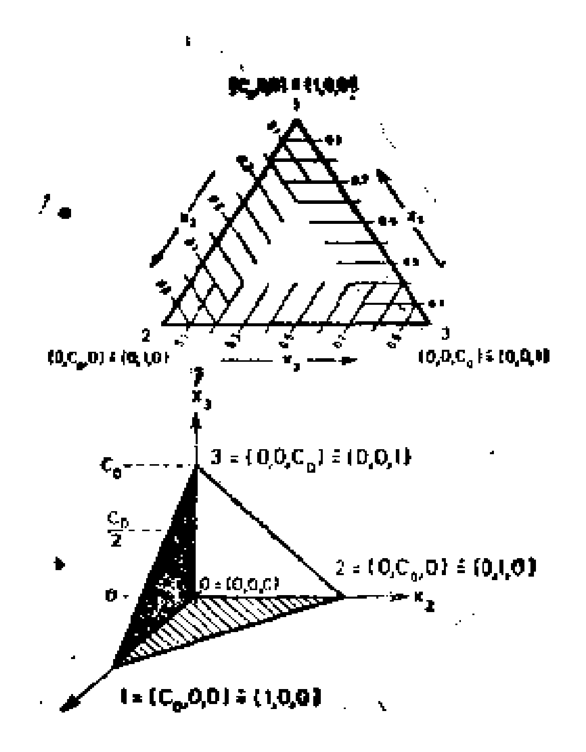


图 1-20 图a 说明了单形S_3，而图b 表明它是怎样被置入物理上可接受的浓度空间的。对于其中一些点，括号内给出了总浓度$c_0=\sum_i x_i$，或者坐标 x_1、x_2、x_3以及 ζ_1、ζ_2、ζ_3。

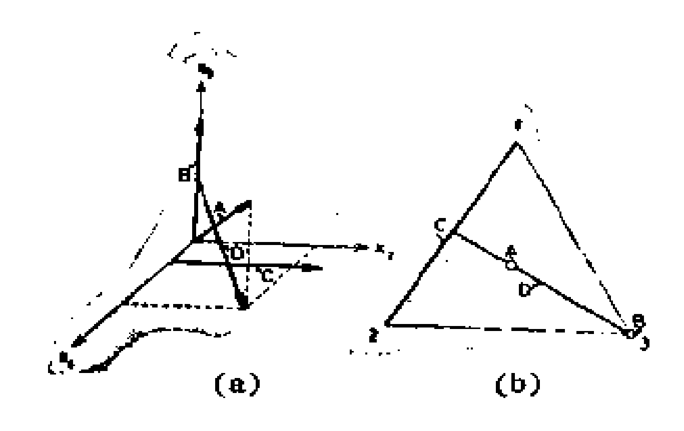


图 1-21 显示某些位点，它们与浓度空间 x_3 (a)及单形 S_3(b)上的总浓度 c_0 有特殊关系。$A=(c_0/3,\ c_0/3,\ c_0/3)$，$B=(0,\ 0,\ c_0)$，$C=(1,\ c_0-1,\ 0)$和$D=(c_0-1,\ c_0-2,\ 2-c_0)$。($a$) 中的箭头标明了随着总浓度增加的点的移动(注意，那些点的所有坐标都象 A 和 B 一样与 c_0 成比例，它们在S_3上描成单点)。

接受的 X^n 的区域则随 c_0 而变动。一般说来，不固定点的位置及其简正模也将依赖于 c_0。为了完整地描述动力学系统的长期行为，有必要估计其自身为总浓度 c_0 的“函数”固定点图，正如我们在后面见到的，许多固定点显示出简单的浓度相关性：它们的坐标与 c_0 成比例。随着总浓度 c_0 的变化，这些点将沿着通过 X^n 的原点的线条移动（参见图 1-21），结果描出了 S_n 上的单点。相应地，固定点图作为一个整体，变得更简单了。固定点图对总浓度值 c_0 的这种形式上的相关，在分析生长系统时具有特殊重要性。

插入n维浓度空间的某一特定的 $(n-1)$ 维超平面的非常对称的那一部分，被称作单位单形。图 1-20 中给出了一个单形的图解，它可以在3维空间加以描述，从而容易形象化。单位单形包括了物理学上有意义的总浓度范围，并且最适合用于图解表示选择过程。

2.3.4 简正模分析法

从一般的非线性微分方程组出发，我们根据 $\dot{x}=0$ 首先确定固定点。对于直接分析一个动力学系统，知道所考虑区域中的所有固定点是重要的。不过，一般说来，这种信息并不是充分的。n 维动力学系统的轨迹，通常将终止于收点。然而，还可能有稳定的封闭轨道或奇异吸引子，通过仔细考察邻域中的固定点的本性并分析向量场，是可以猜出它们的存在的。例如，2维中的稳定极限环可以由彭加勒图来验明。通过简正模分析法，我们能够获得关于固定点本性的信息。

为办到这一点，在给定的固定点 $\overline{X}$ 的邻域中将动力学系统线性化：

$$\dot{z}_i = A_i(\bar{x}) + \sum_{j=1}^{n} A_{ij} z_j + 0(|z|^2) \tag{48}$$

新变量Z_i由下式定义

$$z_i = x_i - \bar{x}_i \quad \text{或} \quad z = x - \bar{x} \tag{49}$$

而系数A_{ij}是一个定义在固定点 $\bar{x}$ 处的$n \times n$的雅各比矩阵(A)中的元素:

$$A_{ij} = \left(\frac{\partial \Delta_i}{\partial x_j}\right)_{x=\bar{x}} \tag{50}$$

由于按固定点的定义有 $A_i(\bar{x}) = 0$，线性化的微分方程组由下式给出

$$\dot{Z} = A \cdot z \tag{51}$$

与简正模有关的时间常数的倒数现在是作为矩阵A的本征值而获得的。本征向量ξ_i决定了相应的浓度变量的线性组合。

$$A \cdot \xi = \omega_j \cdot \xi_j \tag{52}$$

一般地，ω_j 是复量，它们决定着固定点的本性，其中最重要的固定点已总结在图1-19中。

假若矩阵A不是奇异的，那么几乎在所有情况下，线性化系统(51)的一个稳定的固定点都对应于此非线性问题的一个稳定固定点。不过，有一些重要的例外 ($R_e\ w_j = 0$)：此线性系统的一个中心在非线性情况下可作为螺旋收点而出现，并且反之亦然。著名的洛特卡-沃尔特拉模型系统是这种行为的一个例子。我们将遇到另一个例子，即在2.4节中的$n=4$维的超循环。

如果从给定的动力学系统得到一个以上的稳定的固定点、极限环或其他吸引子，那么我们也将希望知道相应的洼域，吸引子对于洼域代表了无限时间处的轨道极限。单个洼域由分界线隔离开，这在原则上可由从鞍点出发沿最陡峭下降的路线的逆向积分($t \to -t$)来决定。如果已知给定动力学系统的所有稳定的固定点和其他吸引子以及它们的洼域，我们就能预见从给定浓度空间中任何一点出发的选择过程的结果。

在某些情况下，我们将有 $R_e\ \omega_j = 0$。于是在不定点附近的线性化不能提供足够的信息，为了完全地描述，人们就必须返回到非线性的动力学系统。通常，直接考察在固定点周围的向量场并不太复杂，因而可得到希望的结果。

确定简正模是固定点分析的内在组成部分。它代表着在固定点的邻域中对动力学系统的轨道进行考察。在大多数情况下，它足以标征出固定点的稳定性质。然而，所涉及到的线性近似往往不足以提供充分的信息，因而往往需要更为高级的分析方法。

2.3.5 生长系统

从(37)式容易导出总浓度 c 的微分方程：

$$\dot{c} = \sum_{j=1}^{n} \Gamma_j(x)\left(1 - \frac{c}{c_0}\right) \tag{53}$$

c_0 代表了由非专一流 ϕ_0 控制的定常总浓度值。

显然，这个方程在 $c=c_0$ 处有一个固定点。其简正模的本征值为：

$$\omega_c = \frac{-1}{c_0}\left(\sum_{j=1}^{n} \Gamma_j(x)\right)c = c_0 \tag{54}$$

只要所有的净生长项 Γ_j 的和为正，这个本征值就为负。从而我们在 $c=c_0$处找到一个稳定的静止态。

在一定的系统中，与内组织有关的不动点图也依赖于总浓度值 c_0。现在，我们应该对于先前的纯数学处理赋予物理学意义了。为做到这一点，我们假定，有一个非平稳动力学系统，它在 $t=t_0$ 处，以具有相应的起始总浓度值 $c(t_0)=c_0$ 开始进化。调节选择约束条件，使得总浓度 $c(t)$ 的变化比动力学系统内部过程的变化慢，这就是说，一切归因于外部过程的变化比归因于内部组织的变化要缓慢得多。在每一瞬间，此系统都趋近一个稳定解（例如一个收点、一个稳定封闭轨道或另一种吸引子）。当前面的条件得到充分满足时，系统就充分接近于长期解，

依赖于时间的过程就可以描述为连续改变总浓度的一系列稳定解。用更为专业化的物理学术语，我们可以说，动力学系统在确定的内平衡条件下发展。正如所期望的，如果内平衡条件得到满足，则对系统的分析可以大为简化。

具有齐次生长函数Γ_i的动力学系统的内平衡是容易分析的，因为此时固定点图S_n并不依赖于总浓度c_0。c_0值从低变到高时，选择行为不会发生任何变化。而且，在生长的齐次系统中，长期的发展也不依赖于内平衡的程度。所以，在这种类型的系统中，选择过程的最终结果将是相同的，而与生长周期内的内平衡是否已确立无关。不过，对于有些情形，若未仔细分析，不可应用内平衡概念。在某些临界总浓度$c=c_{cr}$处，固定点图中可以出现不连续的变化，例如，收点也许变得不稳定，稳定的极限环也许消失了，等等。"霍普夫分歧"是这种不稳定性的一个众所周知的代表。例如象某个生长系统从较低浓度值趋近临界浓度值那样，从某处出发到达了临界点的内平衡动力学系统，在它通过了临界点以后，本质上是要离开平衡态的。在这种点的邻域中对动力学系统进行分析，需格外小心。我们将在第7节中遇到这样的例子。托姆在他的突变论中，已对类似的情形进行了非常全面的研究。

从生物物理学的观点来看，这些复杂的动力学系统当然是更为有趣的。事实上，出现有组织的结构要求剧烈变化，这类似于上述固定点图提到的不连续性。描述不同组织水平之间转移的动力学系统，不可避免地必须通过一定的临界阶段或周期。更具体一些，我们将具体考虑一个例子，它是典型的生物大分子自组织的一个重要问题：从独立竞争者转变成由协同的核苷酸和蛋白质构成的功能单元。按照1.4节的定义，在竞争系统中只有一种物种被选择，因此在这种S_n的内部不存在稳定吸引子。另一方面，任何协同系统都必须具有这种吸引子，否则经过充分长时间后至少有一种参与协同的大分子就会消亡。结果是：一个动力学系统必定在它的一定参数值处包含某种临界不稳定性；这个动力学系统在原则上能够模拟所期望的从比较随机的状态向比较有组织的状态的发展。

2.3.6 分析具体系统

1.独立竞争者

为使得应用固定点分析方法易于理解，作为一个例子，我们考虑在第一部分已讨论过的拟种选择问题。其数学框架在表1-7中列出。浓度空间的坐标由正常变量y_k给出；本征值λ_k是函数Γ_k的生长参数。此分析涉及到突变体的一个给定的分布。对被选择的拟种有所贡献的新突变体的出现，将改变浓度坐标y_k的意义，即改变它们与真实浓度变量x_i的关系。这些结果在表1-7中是不解自明的。我们将利用这些结果，对表1-6中出现的如下3种生长函数$\Gamma_i=k_ix_i^p$进行比较讨论：

(1) 恒生长速率：$p=0$

(2) 线性生长速率：$p=1$

(3) 二次生长速率：$p=2$

表1-7　关于拟种选择的固定点分析(参见第一部分)

速率方程写作：

$$\dot{y}_i=\lambda_i y_i-\frac{y_i}{c_0}\sum_{j=1}^{n}\lambda_j y_j;\ i=1,\ 2,\ \dots,\ n.$$

长期行为由单形S_n的隅角上的n个固定点来决定：

$$\bar{y}_1=\begin{pmatrix}c_0\\0\\\vdots\\0\end{pmatrix},\ \bar{y}_2=\begin{pmatrix}0\\c_0\\\vdots\\0\end{pmatrix},\ \dots,\ \bar{y}_n=\begin{pmatrix}0\\0\\\vdots\\c_0\end{pmatrix}$$

简正模分析对每一个固定点$\bar{y}_k$产生出一个有n个$\omega_j^{(k)}$值的谱系。

$\omega_j^{(1)}=\lambda_j-\lambda_1$	$\omega_j^{(2)}=\lambda_j-\lambda_2$		$\omega_j^{(n)}=\lambda_j-\lambda_n$
$j=2,3,\dots,n-1,n$	$j=1,3,\dots,n-1,n$	…	$j=1,2,\dots,n-2,n-1$
$\omega_0^{(1)}=-\lambda_1$	$\omega_0^{(2)}=-\lambda_2$		$\omega_0^{(n)}=-\lambda_n$

就单形S_n的自由度而论，每一个固定点$\bar{y}_k$都有$n-1$个具有时间常数的倒数$\omega_j^{(k)}$的简正模，$\omega_j^{(k)}$描述不同拟种竞争所造成分布的自组织过程。单形S_n还有一个简正模$\omega_c^{(k)}$，它对应于总浓度c的一个变化。所有内模$\omega_j^{(k)}$由不同的本征值λ来表示。因此，只对最大本征值存在一个稳定的固定点：$\lambda_m>\lambda_j$，$j=1, 2, \ldots, n$，$j\neq m$。它是一个节收点，即所有的$\omega_j^{(m)}$值不相同，且为负。相应地，具最小本征值的拟种由某个发点来描述，因为其ω_j值为正。剩余的$n-2$个固定点则都是鞍点，因为它们含有正的和负的简正模速率常数$\omega_j^{(k)}$。

1．第1种情况产生出一个稳定的固定点，即是单位单形S_n内部的一个焦收点：

$$\bar{x}=\frac{c_0}{\sum_{j=1}^{n}k_j}\begin{pmatrix}k_1\\ k_2\\ \vdots\\ k_n\end{pmatrix} \tag{55}$$

在单位单形"内部"即意味着对$\bar{x}$的所有坐标：$0<\bar{x}_i<c_0$。雅各比矩阵的(负的)本征值是n重简并的：

$$\omega=-\frac{\sum_{j=1}^{n}k_j}{c_0} \tag{56}$$

这对ω_c也成立，ω_c涉及到总浓度c的改变。

结果是所有物种稳定共存。

2．第2种情况在表7中处理。正如我们已知的，在此只存在一个稳定的固定点。它落在单形的隅角上，这个事实标志了竞争行为。节收点的浓度坐标中，只有一个为正($=c_0$)，其余为零。如同第1种情况，此图并不依赖于总浓度c_0，终态结果也不依赖于起始条件。

3．最后，第3种情况显示了总计2^n-1个固定点，这些固定点可以分为3类：

第1类中包括n个焦收点，在S_n的每个隅角上有一个：

$$\bar{x}=\begin{pmatrix}0\\ \vdots\\ x_k=c_0\\ \vdots\\ 0\end{pmatrix} \quad 其中 \quad \begin{aligned}&\omega_j^{(k)}=-k_kc_0\\ &\omega_c^{(k)}=-k_kc_0\\ &j=1, 2, \ldots, n-1\end{aligned} \tag{57}$$

它们是仅有的稳定的固定点。因为是处于单位单形的隅角上，所以它们又标志了一种竞争行为，只允许有一个生存者，即一个纯净态。不过，在这种非线性生长速率的情况下，竞争结果有赖于起始条件，因为存在着n个稳定的固定点（与线性自催化只有一个稳定的固定点形成对照）。这就意味着，n个竞争者中每个都能够以其自己的偏向来决定竞争，这取决于起始群体数。一旦确立了获胜者，其他竞争者就难以生长起来并取代获胜者。

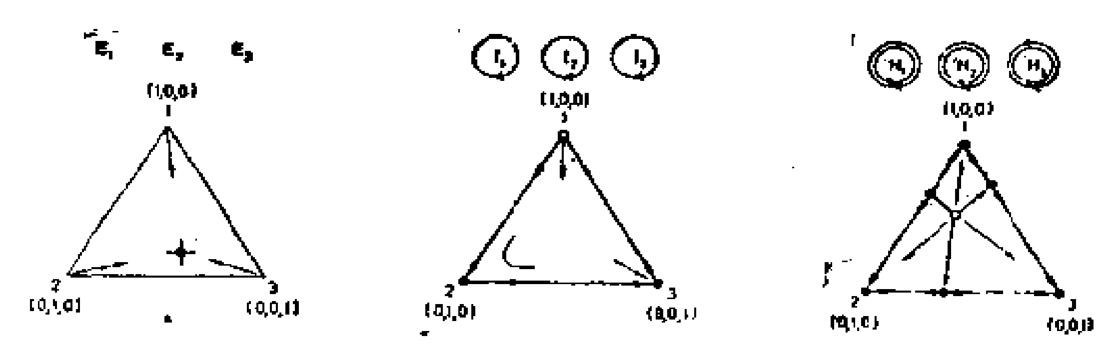


图 1-22 恒组织的不同类型地位竞争者的3维固定点图（符号 E、Ⓘ和◉已在图1-10中引入）。

a）恒生长速率（$p=0$）

$$\dot{x}_i=k_i-\frac{x_i}{C_0}\sum_{j=1}^{3}k_j \quad k_1=1,\ k_2=2,\ k_3=3$$

此图显示出在单位单形S_3内部的一个焦点，这意味着所有3个物种具有稳定的、共存的行为。容易设想轨道的整个流形，这些轨道沿着那些经过单形S_n中每个点的直线而趋近于稳定的焦点。

b）线性生长速率（$p=1$）

$$\dot{x}=k_ix_i-\frac{x_i}{C_0}\sum_{j=1}^{3}k_jx_j \quad k_1=1, k_2=2,\ k_3=3$$

由物种3组成的纯净态，代表了此系统仅有的在长时间范围稳定的解。除了12和13这两条边是例外，其余所有的轨道都始于点1并终于点3。

c）二次生长速率（$p=2$）

$$\dot{x}_i=k_ix_i^2-\frac{x_i}{C_0}\sum_{j=1}^{n}k_jx_j^2 \quad k_1=1,\ k_2=2,\ k_3=3$$

单形S_3被分解为3个区域，每一个区域即是一个稳定的固定点洼域。洼域的大小直接与相应的速率常数值有关。因为k_3是最大的，所以固定点$\bar{x}_3$有最大的洼域。

因此，我们把这种情形称为“一旦则永存的选择”。与前面两种情况一样，此固定点图不依赖于总浓度c_0。

另外两类固定点中包括：单位单形内部的一个发点(所有坐标都是有限的)以及2^n-n-2个鞍点，一个在S_n的每条边上，一个在每个面（包括全部可能的超面)中。这两类固定点都代表了不稳定行为。我们不去列出它们的坐标和简正模，这可由直接计算获得。我们通过一些3维单位单形的例子，来说明不同生长系统的典型的选择行为(参见图1-22)。

上面选用3个比较简单的模型为例，说明了固定点分析方法，并强调了那些我们所关注的性质。固定点的性质，尤其是它提供了稳定解还是提供了不稳定解的性质，对于选择和进化问题最为重要。固定点在单位单形中的位置也同样重要。一组复制单元的协同选择需要固定点落在单位单形S_k的内部，S_k与由这k个协同单元的浓度坐标生成的子空间X^n有关。另一方面，在S_k的一个隅角上，收点的位置代表着竞争，它导致了仅选择一种组分，而在边、面或超面上的位置则表示部分竞争和选择。

例如，构成一个翻译器，要求作为不同基因前体的若干个复制单元的相伴的选择。以上讨论的3种系统都不能满足这种相伴选择的要求。第1个系统看起来是共存的，但它不是选择的，因之不可能向优化功能进化。第2个系统只在拟种分布的非常狭窄的限度之内才允许共存；它不能容忍基因型的趋异，而这是推动表现型的多样化所要求的。最后的第3个系统是强烈的反协同的，其强烈到如此程度，以至一旦确立了一种物种，它就抵制选择任何一种突变体，而不论此种突变体是否提供选择有利性。

按照第5节中比较述评得到的启示，我们现在将深入地研究具有功能联接的系综。这样的系综为了保持遗传信息必须包

含复制单元，同时，还必须通过耦合作用而协同地稳定化，由此使得生长函数内在地是非线性的。这些性质是对连接系统所期望的，因此与第3个独立竞争者的例子有某种关系。

2、催化链

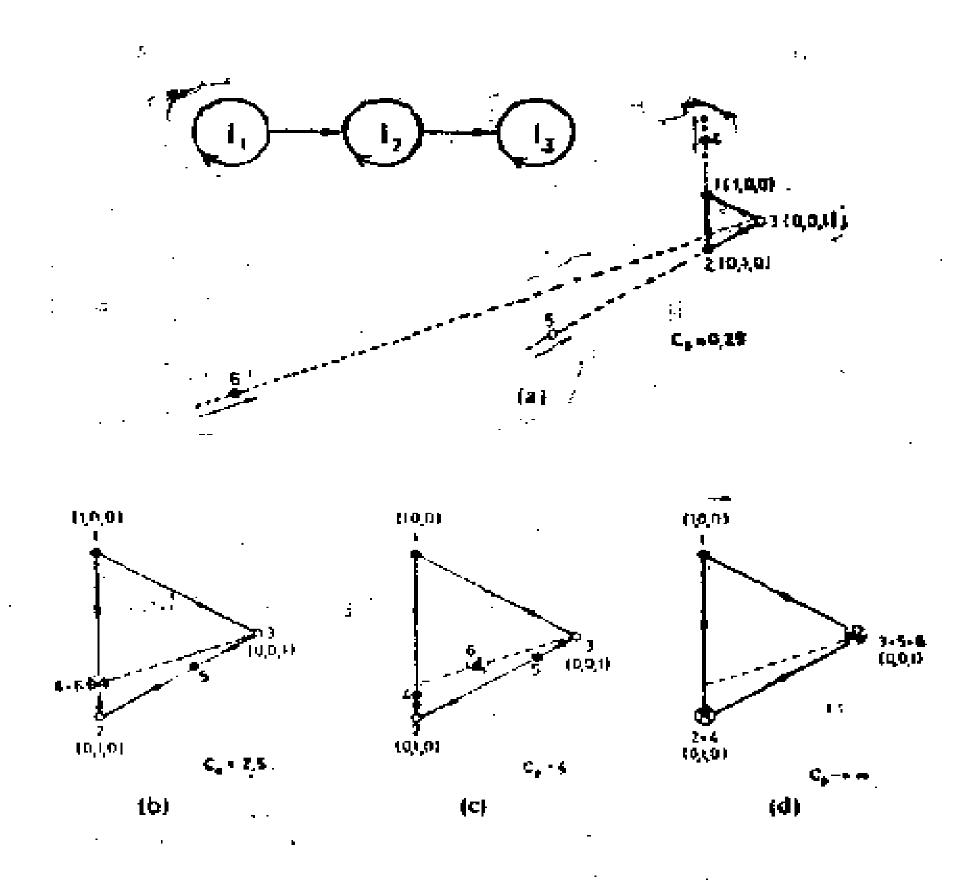


图 1-23　在恒组织约束下的自复制单元催化链的固定点图：

$\Gamma_1 = k_1 x_1$;

$\Gamma_i = k_i x_i + k'_i x_i x_{i-1}$ $(i=2,\ 3)$

$k_1=3$; $k_2=2$; $k_3=1$; $k_2'=2$; $k_3'=1$

$1=\bar{x}_1$; $2=\bar{x}_2, \ldots, 6=\bar{x}_3$

在低浓度(a)，稳定解对应于选择物种1。不过，若总浓度达到某个临界值时其余的两种物种还未区别开来，则出现新的定态，在此所有3个物种都变成稳定的(b)。总浓度(c)的进一步增加，仅仅有利于物种3，所以最后的情形(d)等价于选择物种3。不过，其中的机制不同于独立竞争者的机制。

在一个系综的所有成员之间建立起某种联系的最直接方式，是通过反应耦合而构造起一条链，这很象我们在语言中把词连接成语句(图1-23)。

与这些耦合有关的速率项，将使得净生长函数Γ_i中除开第一个成员以外都是非齐次的：

$$\dot{x}_1 = k_1 x_1 - \frac{x_1}{c_0}\left[k_1 x_1 + \sum_{j=2}^{n}(k_j x_j + k'_j x_j x_{j-1})\right]$$

$$\dot{x}_i = k_i x_i + k'_i x_i x_{i-1} - \frac{x_i}{c_0}\left[k_1 x_1 + \sum_{j=2}^{n}(k_j x_j + k'_j x_j x_{j-1})\right]$$

对于 $i=2, 3, \ldots, n$ (58)

由于没有齐次性，所以，比起前面讨论的固定点图，此类固定点图的形式

表1-8 对3维催化链的固定点分析

固定点图显示了6个固定点，其坐标及简正模如下：		
$\bar{x}_1=\begin{pmatrix} c_0 \\ 0 \\ 0 \end{pmatrix}$	$\bar{x}_2=\begin{pmatrix} 0 \\ c_0 \\ 0 \end{pmatrix}$	$\bar{x}_3=\begin{pmatrix} 0 \\ 0 \\ c_0 \end{pmatrix}$
$\omega_1^{(1)}=k_3-k_1$	$\omega_1^{(2)}=k_1-k_2$	$\omega_1^{(3)}=k_1-k_3$
$\omega_2^{(1)}=k'_2c_0-k_1+k_2$	$\omega_2^{(2)}=k'_3c_0-k_2+k_3$	$\omega_2^{(3)}=k_2-k_3$
$\bar{x}_4\begin{pmatrix} \frac{k_1-k_2}{k'_2} \\ \frac{k'_2c_0-k_1+k_2}{k'_2} \\ 0 \end{pmatrix}$	$\bar{x}_5\begin{pmatrix} 0 \\ \frac{k_2-k}{k'_3} \\ \frac{k'_3c_0-k_2+k_3}{k'_3} \end{pmatrix}$	$\bar{x}_6\begin{pmatrix} \frac{k_1-k_2}{k'_2} \\ \frac{k_1-k_3}{k'_3} \\ k \end{pmatrix}$
$\omega_1^{(4)}=\frac{(k'_2c_0-k_1+k_2)(k_2-k_1)}{k'_2c_0}$	$\omega_1^{(5)}=k_1-k_2$	$k=\frac{k'_2k'_3c_0-k'_3(k_1-k_2)-k'_2(k_1-k_3)}{k'_2k'_3}$
$\omega_2^{(4)}=\frac{1}{k'_2}\{k'_2k'_3c_0-k'_3(k_1-k_2)-k'_2(k_1-k_3)\}$	$\omega_2^{(5)}=\frac{(k'_3c_0-k_2+k_3)(k_3-k_2)}{k'_3c_0}$	$\omega_1^{(6)}$ 和 $\omega_2^{(6)}$ 是雅各比矩阵 $A(\bar{x}=x_6)$ 的本征值

3个固定点$\bar{x}_1$、$\bar{x}_2$和$\bar{x}_3$与单位单形 S_3的隅角一致(参见图1-23)，因之使得竞争行为有意义，而不论竞争行为的本性如何。另外 3 个固定点的位置(线性地)依赖于总浓度c_0。两个固定点$\bar{x}_4$和$\bar{x}_5$沿单形的边$\overline{12}$和$\overline{23}$运动，因此显示了部分竞争。只有固定点$\bar{x}_6$可以通过S_3的内部，产生出所有链成员的协同选择。

在低浓度时：

$c_0<(k_1-k_2)/k'_2$，$(k_2-k_3)/k'_3$或$(k_1-k_3)/k'_3$

固定点$\bar{x}_4$、$\bar{x}_5$或$\bar{x}_6$的位置分别落在单形 S_3 的外部，这意味着落在物理学上有意义的浓度空间区域之外。(至少有一个浓度坐标为负。)对于$c_0\to 0$，这些固定点的位置甚至趋于无穷。此动力学系统渐进地与指数生长的(非耦合的)竞争者系统变得一致，竞争者系统的特征由固定点$\bar{x}_1$、$\bar{x}_2$和 $\bar{x}_3$ 标志。

若$k_1>k_2$，k_3和 c_0在由$[(k_1-k_2)/k'_2]+[(k_1-k_3)/k'_3]$给出的阈值之上，则标志协同行为的固定点 $\bar{x}_6$进入单位单形之中。不过，它并不趋近 S_3 内部的任何点，而是迁向隅角3。

更复杂。

为保持这种程序清晰易懂，我们从 3 维系统出发，然后再把我们的分析推广到更高维的系统。表1-8把第 3 维的相干关系以及固定点图的简明特征汇集在一起。依据这种分析，自复制单元的线性链上的3个成员(I_1 至 I_3)，只有在如下非常特殊的条件下才能被相伴地选择：

$$k_1>k_2,\ k_3 \tag{59}$$

以及

$$c_0>\frac{k_1-k_2}{k'_2}+\frac{k_1-k_3}{k'_3} \tag{60}$$

看来极不可能的是，碰巧满足条件(59) 的这些成员能在很长的进化期间中保持这种关系[这意味着决不会出现改变关系(59)的突变]。即使它们能做到这一点，此系统也将以高度不对称的方式发展，从而至少在选择约束下，只有链中的最后一位成员的群体数会随c_0 而增加。注意到这是马上意味着群体数以若干个数量级发散，我们可以得出结论，认为这种系统不可能稳定一个联合函数，因为它不可能在大的总浓度范围控制有关的群体数。

这种行为在图1-23的一些例子中得到了说明，它们对于在接近内平衡

阶段的生长系统的连续过程，提供了一些瞬时图象。对于低于(60)式给出的临界总浓度c_0，3个固定点$\bar{x}_4$、$\bar{x}_5$和$\bar{x}_6$落在单位单形的外部(图1-23a)。如果c_0等于临界值，则固定点$\bar{x}_6$到达此单形的边界（图1-23b)，并且随c_0的增加而迁移通过其内部。在此同时，它的本性也已改变，它现在代表了一个稳定固定点(图1-23c)，在此例中是一个螺线收点。图1-23b 表明了这个稳定固定点的结局，亦即它迁到隅角3。因之此系统趋近纯净态$\bar{x}_3=c_0$。

对3维情况所获的有关结果，容易推广到n维系统。物种3的地位现在由物种n所取得。代替6个固定点，我们找到的是$2n$个固定点。最吸引人的固定点是$\bar{x}_{2n}$。容易确定它的位置：

$$\bar{x}_{2n}=\begin{pmatrix} \dfrac{k_1-k_2}{k_2'} \\ \dfrac{k_1-k_3}{k_3'} \\ \vdots \\ c_0-\sum\limits_{j=2}^{n}\dfrac{k_1-k_j}{k_j'} \end{pmatrix} \tag{61}$$

当且仅当速度常数满足关系$k_1>k_j$，$j=2, 3, \dots, n$，并且总浓度超过如下的临界值：

$$c_{cr}=\sum_{j=2}^{n}\frac{k_1-k_j}{k_j'} \tag{62}$$

此时，固定点$\bar{x}_{2n}$才会落在单形S_n的内部。因此，$\bar{x}_{2n}$相应于一个稳定的定态。除了$\bar{x}_n$以外的所有浓度在此状态中都是常量，因此在大的总浓度时此系统趋近于纯净态$\bar{x}_n=c_0$。

我们可以对催化链的行为小结如下：

①　只有当速率常数和总浓度满足如下关系时，稳定的静止态才存在：

$$k_1>k_j,\ j=2, 3, \dots, n;\ c_0>\sum_{j=2}^{n}\frac{k_1-k_j}{k_j'}$$

为了抵制其他的非功能单元而进行选择时，此系统必须遵守选

择约束条件，并且在选择有利突变体时不能改变所要求的速率常数的量级。

（2）假若在(1)中的条件得到了满足，则只有在相当小的总浓度范围内单个物种的浓度才有类似的大小。随着c_0值增加，链上的末位成员I_n在(准)稳定条件下生长，最终将唯一地处于支配地位。

所以，催化链不象作为信息整合系统那样有效。

3. 分支系统

无论耦合系统在何处进化，将不可避免出现如图1-24所描绘的耦合分支。对这种分支网络进行固定点分析，不会揭示任何意外的新特征。在总浓度非常低时，3种行为就象独立的竞争者。现在c_0有两个临界值，在临界值处I_1和I_2或I_1和I_3成为共存。这最终取决于是I_2还是I_3有效地取得I_1的偏爱。两个固定点中结果一个是稳定的节点，另一个是鞍点。在总浓度较高时，稳定固定点又分别迁到隅角2或隅角3。图1-24已对这种行为进行了说明。

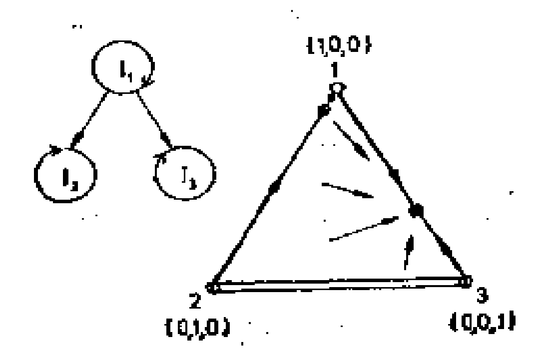


图 1-24 某动力学系统的固定点图 在恒组织约束下的自复制单元(I_i)的催化网络中出现了分支点

$\Gamma_1=k_1x_1$; $\Gamma_i=k_ix_i+k_2'x_ix_1$ (对于 $i=2,3$)

$k_1=3$; $k_2=k_3=0$; $k'_2=1$; $k'_3=2$; $c_0=3.5$

这里所研究的3维系统，能够以两种方式加以推广：

1. 可以从一个给定点发出两条以上的分支。

2. 单条分支中可以含有一个以上的成员。对这些多维系统进行固定点分析所得的结果，本质上与从3维系统中获得的那些结果是相同的，并可以小结如下：

自复制单元分支系统不会长时间保持稳定。在生长中很有效的分支将被选择，而其余的分支将消失。最后所剩下的是最有效的线性链，因此，整个问题被归结为类型(58)的动力学系统，我们已在前面的章节中讨论过这类系统。

2.3.7 超循环的固定点分析

1. 分类

正如我们在第一部分中见到的，动力学系统中环的闭合导致了此种系统作为一个整体具有全新的性质。在化学反应封闭环中形成的分子集合代表了一种催化剂。催化剂的循环(图1-4)反过来又具有自催化性质，可能被当作自复制系统。在发现了自复制单元之间直接的或者分支的耦合并不导致功能连接系统的联合选择以后，我们可以问，是否在耦合链中任何环的闭合都将改变总系综的选择行为的本性。我们也许会争辩道，情况必定如此，因为我们记得在开链中末位成员总是得到耦合带来的全部好处。

在第一部分中，我们已经一般地对超循环进行了分类。这类网络的最简单成分由自复制单元之间的直接功能连接而获得，如图1-25所示。

本节对超循环的讨论由3个不同的部分组成。首先，我们对于这类新的催化系统的分类提出了一些定义和有用的标准。

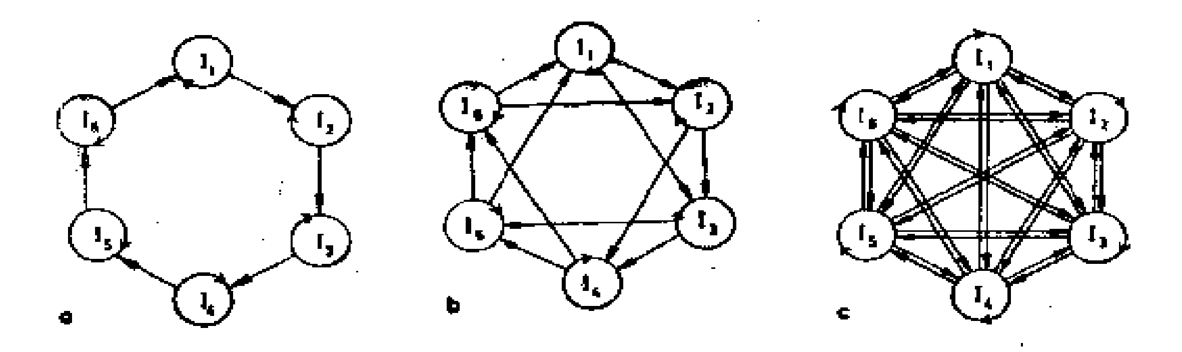


图 1-25 催化超循环 a）生长度 $p=2$，$n=6$；b）生长度 $p=3$，$n=6$；c）生长度 $p=n=6$。

然后描述了最重要的"纯净"型超循环固定点分析的结果。最后，我们将研究自组织系统的一个例子，这种系统代表了一种现实的催化超循环。

超循环不同于普通循环，最基本之处在于生长速率中的非线性项。在简单情形中，如(63)式所示，函数 Γ_i 将是浓度的产物。下式中：

$$\Gamma_i = k_i \prod_{\lambda=1}^{n} x^{p_{\lambda i}} \tag{63}$$

指数 $p_{\lambda i}$ 可以被称作矩阵 P 的元素。下标 λ 和 i 表明，在函数 Γ_i 中是哪一个群体变量 (x_λ) 必须自乘到 $p_{\lambda i}$。于是动力学系统完全由指数矩阵 P、速率常数向量 $k=\{k_1,\dots,k_n\}$ 以及一组初始条件所决定。首先，我们将只研究"纯"情况，它们是以齐次生长项 Γ_i 为特征的。齐次性要求导致了对于矩阵 P 中指数的第一个限制：

$$\sum_{\lambda=1}^{n} p_{\lambda i} = p,\quad i=1,\ 2,\dots,\ n \tag{64}$$

现在"p"为所有的 n 个微分方程所共有，它代表了在第 5 节中引入的生长函数的生长度。加上齐次性的限制，我们将允许个体浓度在 Γ_i 中仅以一次项出现。一些具有比较高级相关的例子将在后面讨论。相应地，指数 $p_{\lambda i}$ 被限制到只有两个可能的值：$p_{\lambda i}=\{0,\ 1\}$。

最后让我们把循环对称引入到净生长函数中：

$$\Gamma_i = k_i x_i x_j x_k x_l \ldots x_r$$

$$j = i - 1 + n(\delta_{i1});\quad l = i - 3 + n(\delta_{i1} + \delta_{i2} + \delta_{i3});\ \ldots$$

$$k = i - 2 + n(\delta_{i1} + \delta_{i2});\quad r = i - p + n\sum_{\mu=1}^{p-1}\delta_{i\mu} \tag{65}$$

对于解的最重要特征来说，循环对称的假设并不是最基本的。不过，如果在此循环系统中，关于各个成员的动力学方程中在结构差异方面没有更多信息可资利用，那么它就是一个合理的假设。矩阵P现在具有了一般的、简单的形式。作为一个具体的例子，$n=6$和$p=3$的矩阵P为：

$$P(n=6,\ p=3) = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \tag{66}$$

因此，具有循环对称和齐次生长函数Γ_i的超循环，由值n和p以及向量k完全确定。

图1–25是$n=6$且$p=2$、3和n超循环的三种情况的模式图。必须从称之为超循环的这一类催化系统中排除掉$p=1$的情况，因为它们被划入具有线性生长速率Γ_i的系统族。

2．一般分析

表1–9是对超循环系统进行固定点分析的小结。

我们讨论本文中十分重要的两个情形：

(1) 最简单的超循环，可能是$p=2$；

(2) 利用所有成员之中的催化联接的超循环，即对于$i=1,\cdots,n$，$p_{\lambda i}=1$，以及$\lambda=1,\cdots,n$，从而$p=n$。

对第一种系统，我们简单地把它命名为“基本超循环”，对第二种系统，按照其作为协同行动的复合体最为共同的物理状

表1-9　超循环的固定点图

使动力学系统(65)服从恒组织条件，我们得到：

$$\dot{x}_i = k_i x_i x_j \ldots x_l - \frac{x_i}{c_0} \sum_{r=1}^{n} k_r x_r x_s \ldots x_t$$

$$j = i - 1 + n\delta_{i1},\ \ldots,\ l = i - p + 1 + n \sum_{\mu=1}^{p-1} \delta_{i\mu}$$

$$s = r - 1 + n\delta_{r1},\ \ldots,\ t = r - p + 1 + n \sum_{\mu=1}^{p-1} \delta_{r\mu};\ p \leqslant n \tag{T9.1}$$

对于任一 n 值，如果所有速率常数[a]相等，

$$k_1 = k_2 = \ldots = k_n = k \tag{T9.2}$$

则可以用分析手段对固定点进行分析。所获结果如下：

1）用$\bar{x}_0$标记的一个固定点，总是位于浓度单形的中心。

2）n个其他固定点 $\bar{x}_1$、$\bar{x}_2$、…、$\bar{x}_n$，出现在单形 S_n 的隅角。

3）在许多情况下，还有1、2和3甚至更高维的固定点流形，例如固定点边、三角形、四面体和更高维的单形。这些流形总是出现在相应单形S_n的边界上，例如对 $n \geqslant 4$，在 S_n 的边界上有固定点边；对 $n \geqslant 6$，在S_n的边界上有固定点三角形；对 $n \geqslant 8$，在 S_n 的边界上有四面体。

在与协同选择相应的中心固定点 $\bar{x}_0$ 附近所进行的简正模分析得出：

$$\bar{x}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \frac{c_0}{n};\quad \omega_j^{(0)} = -\frac{1-\gamma^{-(p-1)}}{1-\gamma} \left(\frac{c_0}{n}\right)^{p-1} \cdot k$$

$$j = 1,\ 2,\ \ldots,\ n-1;\quad \gamma = e^{\frac{2\pi i}{n} j},$$

$$\omega_n^{(0)} = \omega_c^{(0)} = -\left(\frac{c_0}{n}\right)^{p-1} \cdot k \tag{T9.3}$$

对于 $p=2$，有n个或 $n-1$个不同的本征值，而对于 $p=n$，则所有的本征值都等于$\omega_c^{(0)}$。第一种情况是通常的情形，对于 $p=2$ 和偶数n，获得 $n-1$ 个单重简并的本征值和一个双重简并的本征值 $\omega_c^{(0)} = \omega_j^{(0)} = n/2 = -k(c_0/n)^{p-1}$，而对于奇数 n，则所有的本征值都是相异的。$\omega_c^{(0)}$ 的负值又表明，在单形 S_n 上动力学系统在总浓度 c 涨落时是稳定的。

---

[a]. 关于这些解的个体速率常数的变分，将在第8节讨论。

况，我们将其命名为“复合超循环”。

3．基本超循环

依据动力学系统的维数，我们观察处于单形中心的固定点本性的有趣变化。为此目的，我们更仔细地审查对不同n值所获的一组本征值，它们被恰当地描述为在高斯复平面上的向量$\omega=R_e\omega e_1+iI_m\omega e_2$（图1-26）。处于中心的固定点，对于$n=2$是焦

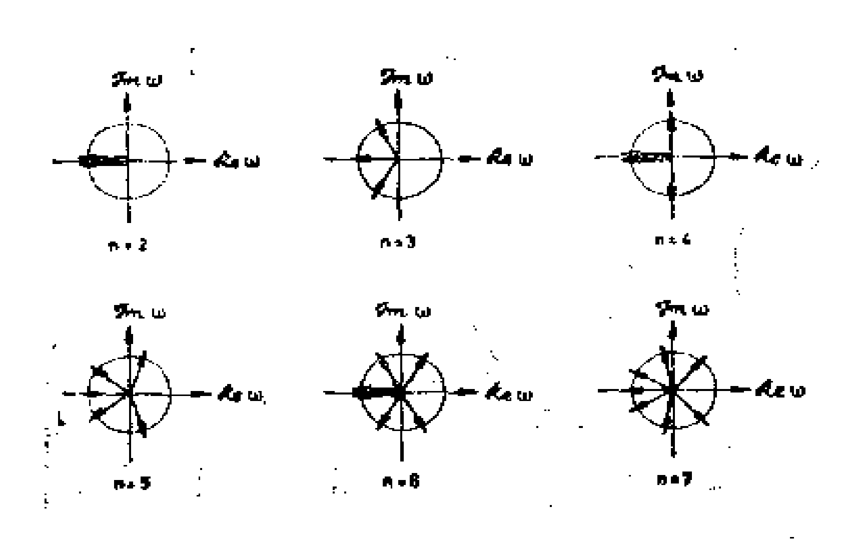


图 1-26　$p=2$、维数为n的(65)型超循环的中心不动点的标准模ω。$R_e\omega$和$I_m\omega$分别代表了频率ω的实部和虚部。

点，对于$n=3$是螺线收点，对于$n=4$是一中心。对于$n\geqslant5$，在一些平面中我们得到最有螺旋分量的鞍点。固定点本性中的这些特征变化，使我们联想起霍普夫分支，尽管我们的参数是离散变化的量，即动力学系统的维数n。正如第8节的更一般分析将表明的，对于$n=2$、3和4，此中心固定点是渐近稳定的。在更高的维数$n\geqslant5$的情况，我们发现了一个更复杂的吸引子，即一个稳定的封闭轨道或极限环，它总是处于单形内部而从不会到达边界。在后一种情况下，浓度x_i的时间平均：

$$X_i(t)=\frac{1}{t}\int_0^t x_i(\tau)d\tau,\quad \overline{X}_i=\lim_{t\to\alpha} x_i(t) \tag{67}$$

将迅速地趋近c_0/n(对于相等的k值)，这正好与稳定固定点的情况有相同的值。

对于处于S_n的任一隅角$k(x_k=c_0)$的固定点，我们发现1个为正的、$n-1$个为零的$\omega_j^{(k)}$值。在第8节中，我们将分析这些非线性贡献，并证明这些固定点是鞍点。因此，相应的长期解对选择行为没有贡献。

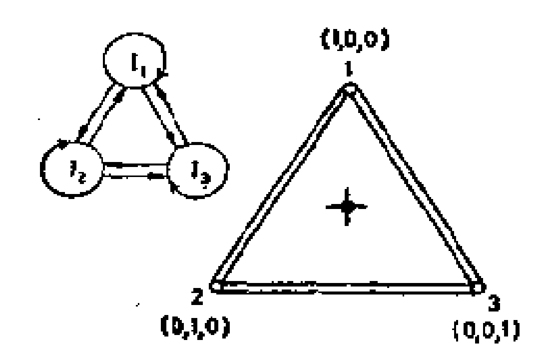


图 1-27 在恒组织约束条件下，由自复制单元①所构成的动力学系统(69)的固定点图，这些自复制单元形成一个超循环。
[$\Gamma_i=k_ix_ix_j$; $j=i-1+n\delta_{i1}$; $n=3$, $p=2$, $k=(1,1,1)$]

作为一个具体的例子，图1-27显示了$p=2$且$n=3$的一个超循环的固定点图。

一般来说，形成一个超循环动力学系统的各个自复制单元的净生长函数，不仅含有催化项，还将含有一级生长项:

$$\Gamma_i=k_ix_i+k_i'x_ix_j \tag{68}$$

使具有这些生长函数的动力学系统服从于恒组织约束，我们得到:

$$\dot{x}_i=k_ix_i+k_i'x_ix_j-\frac{x_i}{c_0}\sum_k(k_kx_k+k_k'x_kx_l) \tag{69}$$

$j=i-1+n\delta_{i1}$，$l=k-1+n\delta_{k1}$; $i=1, 2, \dots, n$

从数学的观点来看，催化链(图1-23)不同于超循环正好在于一个速率常数的不同，取$k'_i=0$，由后者得到催化链。因此我们预期这两种类型的动力学系统之间有些相似之处。与生长函数的非齐次性一致，固定点图依赖于总浓度c_0处于低浓度极限时，两个系统变得与指数生长的独立竞争者系统一致(图1-22)。另一方面，处在高浓度时，这两个系统则不一样。动力学系统(69)渐进地类似于相应的基本超循环($p=3$)动力学系统。

作为一个实例，我们再一次考虑一个维数$n=3$的系统。那里有7个固定点：其中3个与单形S_3的隅角重合，另外3个固定点落在边上，第7个点则处于S_3的内部。

对给定的一组参数k进行计算，其数字结果显示在图1-28中。正如对图1-23的催化链一样，我们提出3个不同总浓度c_0的固定点图，分别代表了低的和高的浓度极限(a和c)以及临界点(b，$c_0=c_{cr}=k_3(k_1'^{-1}+k_2'^{-1})-k_1k_1'^{-1}-k_2k_2'^{-1}$)。

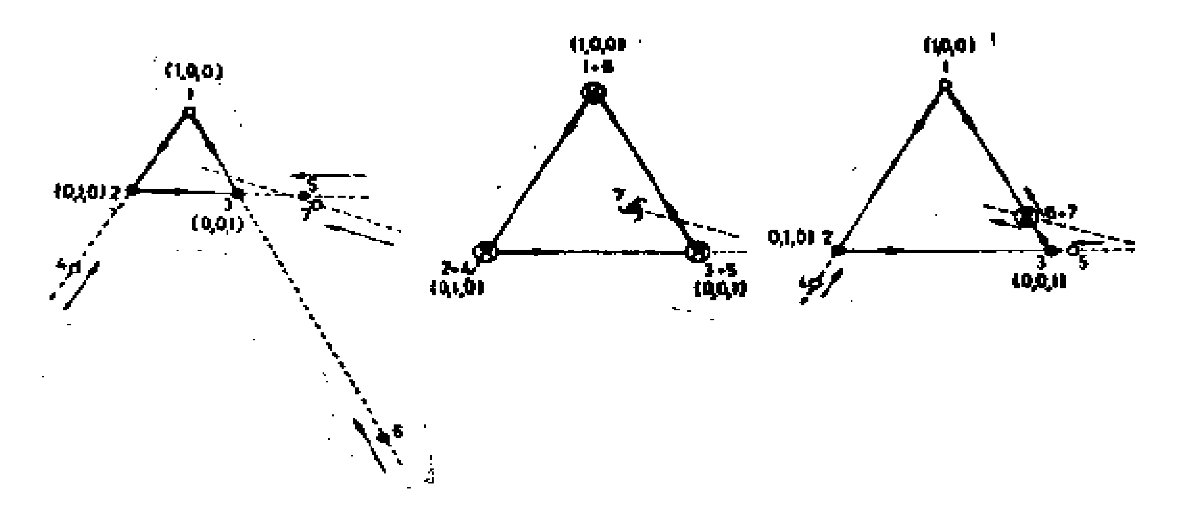


图 1-28 由自复制单元所构成的动力学系统(69)的固定点图，这些自复制单元形成催化超循环。
($\Gamma_i=k_ix_i+k'_ix_ix_j$，$j=i-1+n\delta_{i1}$；
$n=3$，$p=2$，$k=(1,2,3;1,2,3)$；
a) $c_0=0.5$，b) $c_0=2.5$，c) $\lim c_0\to\infty$；
$1=\bar{x}_1$，$2=\bar{x}_2$，$3=\bar{x}_3$，$4=\bar{x}_{12}$，
$5=\bar{x}_{23}$，$6=\bar{x}_{31}$，$7=\bar{x}_0$)

考虑动力学系统(58)和(69)在接近内平衡时的发展，我们认识到在循环系统和非循环系统之间的一个非常重要的差别：循环系统导致了某个渐进的高浓度极限，此极限以各个物种的恒定相对浓度为特征，而开放链在高的总浓度处则趋近纯净态$(x_n = c_0)$。

对从低浓度极限到高浓度极限的整个发展进行总结，使我们认识到，正如动力学系统(69)所描述的，一个超循环是自组织的一个合适的例子。从个体物种之间的竞争出发，此生长系统趋近一个终态，在此状态下所有成员的纯产量是受动态控制的。这种内部控制导致了一个稳定的定态，或者导致其群体变量在固定点附近作规则振荡的状态。

4．复合超循环

对$p=n$的情况进行分析，导致了一个简单的、一般的结果：如同在前面所讨论的其他例子一样，在单形内部有一个固定点。不过，单形的整个边界是由不稳定的固定点、固定点边和固定点面等构成的。因为，对于所有的n值，其中间$(\bar{x}_0)$的不变点是焦点，所以，从整个都具有物理意义的区域——单形内部出发的所有轨迹，在充分长的时间后都将趋近这个特定的点。伴

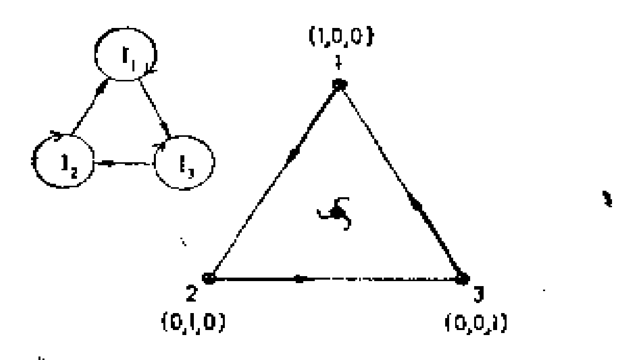


图 1-29 在恒组织约束下形成复合超循环的自复制单元I_i所构成的动力学系统(65)的固定点图。

$[\Gamma_i = k_i x_1 x_2 \ldots x_n;\ n=3,\ p=3,\ k=(1,1,1)]$

随 $\bar{X}_0$ 的所有本征值 $\omega_i^{(j)}$，对于给定的k、c_0和n的值都是相同的。如果我们使$p=n$，则它们都遵从表1-9中的表达式(3)。图1-29显示了一个其$n=3$的复合超循环的固定点图。所以，这种复合体对于控制其组成的相对浓度，是一个极好的例子。

5. 各种超循环的比较

超循环具备内在的整合信息的能力。的确，其中最简单的超循环代表了复杂程度最低的动力学结构，这种结构能使功能上相联的自复制单元的系综免受由于选择竞争消灭了它们的一些成员而导致的信息丧失。从动力学观点来看，所有各种超循环在这种性质方面都是等价的。另一方面，复杂程度较低的系统，诸如简单的催化循环(图1-4)，因为它们缺乏内在的自复制性，所以不配作为整合信息的系统。

根据自然界的现实性，可进一步对超循环的等级进行区分，这是第三部分的主题。为提出这种论据的一个例子，我们在此把简单的超循环($p=2$)与复杂的超循环($p=n$)，就它们的物理学实际方面加以比较。简单的超循环，依据其生长规律，要求生物分子的大分子碰撞。这些生物分子项容易由各种机制提供，同时也是来源于关于核酸复制或者信使指令的蛋白质合成的现实假定(参见2.5节和第三部分)。复合的超循环要求所有的伙伴都对每一组分的生成速率有所贡献。于是导致这种复合超循环的机制将需要(极不可能的)多分子碰撞或是形成由n个不同的亚单元组成的中间物(这在低浓度时是极难形成的)。另一方面，前生物条件的特征是一些特殊的大分子的浓度极低。为了通过复合超循环有效地开始进化，人们将不得不假定远远超出实验经验以外的极高的结合常数，以及在这些常数与单种组分的功能效率之间的内在联系。因此，对于翻译系统的集成过程来说，

与任何低组织度 p 的超循环相比，复合超循环很可能不那么重要。在前细胞进化的更高级阶段，复合超循环也许已获得了生成的机会。

由具有催化活性的和自复制的单元构成的各种系统，已经由固定点分析加以研究。结果清楚地表明了超循环耦合的必要性。只有催化的超循环能够满足在第2.45节中列出的信息整合的标准。

1. 每一组分的选择稳定性是由于与错误拷贝进行有利的竞争。

2. 各组分的协同行为整合入新的功能单元。

3. 此功能单元与其他一些效率较低的系统进行有利的竞争。

2.4 基本超循环的动力学

超循环作为前生物自组织的有关系统，值得对它们的动力学行为进行更详细的分析。对于此类基本超循环($p=2$)，直至维数为4，都可以给出完整的定性描述。对于更高的维数，以及对于结构更复杂的超循环，我们在进行拓扑分析时不得不借助于数值积分。我们以基本超循环为例说明这种方法，这种方法揭示了超循环自组织的全部基本性质①。

① 作为超循环的特例，其中生长函数的2级项 $x_k x_{k-1}$ 已被 $x_k \ln x_{k-1}$ 形式的项取代了，可以获得解析解。

2.4.1 定性分析

由于我们关心的是各组分协同作用的动力学系统，所以最关注的是在物理上可接受的浓度范围内的稳定吸引子。更具体说来，我们必须研究那些固定点的稳定性，对于它们，雅各比矩阵的一些本征值的实部为零。在第7节中(表1-9)，我们主要遇到两种情况:

1．对于单形S_n的隅角上的固定点$\bar{x}_i$，本征值为零($\omega_j^{(i)}=0, j=2, 3, \ldots, n, i=1, 2, \ldots, n$)。

2．单形S_4^2上的4元超循环的中心固定点，是纯虚数本征值($\omega_{2,4}^{(0)}=\pm i$)。①

在对于低维数$n\leqslant 4$的超循环的中心固定点的稳定性作出一般的证明之前，让我们首先详细检查一下这些系统的拓扑性。

相应于基本超循环的动力学系统可以分解成几个子系统，在整体上不变的子空间定义每一个子系统。当且仅当通过子空间任一点的轨迹从不离开子空间时，一个点集合或者子空间将被称作关于给定动力学系统的“整体不变量”。

特别是我们发现，单形S_n上的动力学系统可以划分为在边界上的单形(BS_n)和在内域的单形(IS_n)。单形的内域与以前一样定义，它是群体变量不会消失的区域，即$0<\zeta_i<1$，$i=1, 2, \ldots, n$。显然，我们最感兴趣的是IS_n上的动力学系统，因为它们描述了完整的超循环的发展。我们在下面以数字2，3，4，…，N来标记它们。

在边界上，有1个、2个以至更多个群体变量消失了。结果是，在BS_n上的动力学系统可以划分为诸如边、面和超面那样的低维数的单形上的动力学系统。为把这些系统与完全的超循环加以区别，我们将利用2*A*、2*B*、3*A*等作为速记符号。所有对应于维数$n\leqslant 4$的基本超循环，都示意在图

① 在$4k$(k是不小于2的整数)维的基本超循环中，也将出现纯虚数的本征值，不过，在这些更高维的例子中，中心不动点是鞍点，与对纯虚本征值的更高级贡献的性质无关。

1-30中。表1-10给出了把4维的系统分解成为 11 个子系统的具体例子。

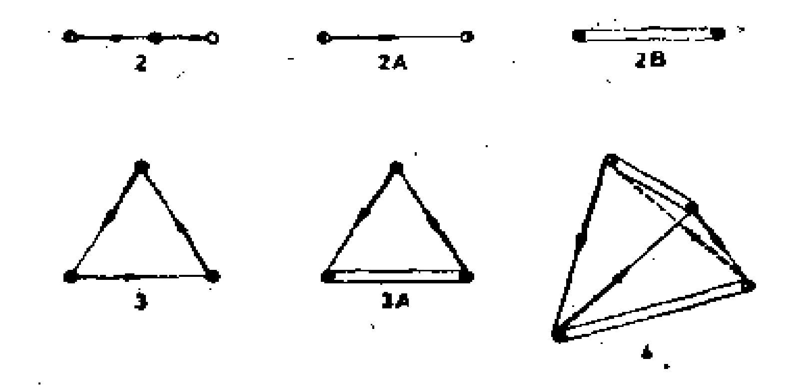


图 1-30　对应于维数$n=2$、3和4的基本超循环的动力学系统。单形S_n上图示的各个系统，可以分解成为整体上不变的动力学子系统(表1-10)："完全的"子系统 2、3 和 4，由所有非零群体变量来标征，从而描述了在单形S_n的内域中(IS_n:$0>x_i>c_0, i=1, 2, \dots, n$)的发展。在单形$S_n$的边界上($BS_n$)，1个或更多个群体变量消失了，同时获得了低维动力学子系统，诸如"流动边"2A、"固定点边"2B、3A 型三角形(注意：动力学系统 2A 出现在3、3A 和 4 的边界邻域中，2B 出现在 3A 和 4 的边界邻域中，3A 最后出现在 4 的边界邻域中)。

表1-10　维数 $n=4$ 的基本超循环，其动力学子系统在整体上不变

符号	条　　件	动力学系统
4	$\zeta_1,\ \zeta_2,\ \zeta_3,\ \zeta_4>0$	$\dot\zeta_i=\zeta_i\zeta_j-\zeta_i\phi,\ i=1,\ 2,\ 3,\ 4$ $j=i-1+n\delta_{i1}$ $\phi=\zeta_1\zeta_2+\zeta_2\zeta_3+\zeta_3\zeta_4+\zeta_4\zeta_1$
3A	$\zeta_4=0$ $\zeta_3=0,\ \zeta_2=0$ 或 $\zeta_1=0$	$\dot\zeta_1=-\zeta_1\phi$ $\dot\zeta_i=\zeta_i\zeta_j-\zeta_i\phi,\ i=2,\ 3,\ j=i-1,$ $\phi=\zeta_1\zeta_2+\zeta_2\zeta_3$ 类似地
2A	$\zeta_3=\zeta_4=0$ $\zeta_2=\zeta_3=0,\ \zeta_1=\zeta_2=0$ 或 $\zeta_4=\zeta_1=0$	$\dot\zeta_1=-\zeta_1\phi,\ \phi=\zeta_1\zeta_2$ $\dot\zeta_2=\zeta_1\zeta_2-\zeta_2\phi$ 类似地

续表

符号	条　　件	动力学系统
2B	$\zeta_2=\zeta_4=0$ $\zeta_1=\zeta_3=0$	$\zeta_i=0,\ i=1,2,3,4$

到维数 $n=4$ 为止的所有系统，都可以用李雅普诺夫方法加以分析(表1-11)。对于2、3和4这3个系统，我们给出了李雅普诺夫函数，因此中心固定点代表了一个稳定吸引子。而且，这个固定点的洼域扩展为单形的整个内域。用更专门的物理学术语来说，这意味着：从任何群体变量分布出发，我们都终止在同一个稳定的定常浓度的集合。动力学系统的确是以各组分的协同行为作为特征的。这个结果对于4维系统具有特殊的重要性，在此，固定点分析法中所运用的线性近似产生出一个被 x、y 平面上的同心封闭轨道流形所包围的中心（见图 1-31a），这无法对稳定性作出明确的结论。

在单形边界上的动力学系统(BS_n)，决定了"破缺的"超循环的行为，"破缺的"超循环即其中至少缺少一个成员的催化超循环。实际上，这些系统描述了超循环消失的动力学。它们在超循环的生成阶段也具有某种重要性。到维数为4为止的完整的动力学系统的边界上，我们发现了 $2A$ 和$2B$两种边以及 $3A$ 面（图1-30）。所有这3个动力学系统都可以直接加以分析。第一种边 $2A$ 联系着两个顺序相连的纯净态或隅角，它们标以"i"和"j"（$j=i+1-n\delta_{in}$）。正如图 1-32 中表明的，沿着边存在一种稳定的推动力，指向总是 $i\to j$。因此这个系统的唯一轨道是从隅角 i 走向隅角j。相应地，我们将把系统 $2A$ 称作流动边。在趋近隅角 j 中，推动力以抛物线方式减少(图 1-32)。因此，泰勒展开式的线性项在固定点 $\bar{x}_j$ 消失，固定点分析无法就这个固定点的性质作出预见。

表1-11　维数$n=2$、3和4的基本超循环的李雅普诺夫函数

为证明动力学系统 $\dot{x}=V(x)$的一个固定点的稳定性，我们必须找出一个任意函数 $V(x)$，它满足如下的两个标准：

(1) $V(\bar{x})=0$　且 $V(x)>0, x\varepsilon U$　　(T11.1)

即在固定点处函数消失，而在不动点的邻域 U 中函数为正。所以，$V(x)$在不动点处具有局部的极小值。

(2) $$\dot{V}(x)=\frac{dV}{dt}=\sum_{i=1}^{n}\left(\frac{\partial V}{\partial x_i}\right)\frac{dx_i}{dt}<0,\quad x\varepsilon U \qquad (T11.2)$$

即在固定点的邻域中$V(x)$的时间导数为负。$\dot{V}$在$\bar{x}$处消失的原因很普通：$\dot{V}(\bar{x})=0$。如果对给定的动力学系统的固定点能够找到这样的函数 $V(x)$，它就称作一个(严格的)李雅普诺夫函数。此不动点$\bar{x}$是(渐近地)稳定的：所有通过$\bar{x}$的邻域中一点的轨道都将终止于固定点 $\bar{x}$。

也可以定义一个较不严格的李雅普诺夫函数 $V(x)$，它满足较弱的标准。

$$\dot{V}(x)\leqslant 0 \qquad (T11.3)$$

任一进入$\bar{x}$的邻域 U 的轨道都将保留在那里。一个具体的例子是：在较强的李雅普诺夫意义上收点是渐进稳定的，而中心只是关于较弱的标准（T11.3）是稳定的。

为方便起见，当我们把李雅普诺夫方法用于基本超循环时，我们使用归一化的变量 ζ_i，并假定速率常数都等于 1，即 $k_1=k_2=\ldots=k_n=1$。

函数

$$V=\left(\frac{1}{n}\right)^n-\zeta_1\zeta_2\ldots\zeta_n \qquad (T11.4)$$

有一个极小值，并在固定点$\zeta_i=\frac{1}{n}$处消失，所以满足条件(T11.1)。V的时间导数可由直接求导得到

$$\dot{V}=-\zeta_1\zeta_2\ldots\zeta_n(1-nr);$$

$$r=\sum_{k=1}^{n}\zeta_k\zeta_l;\quad l=k-1+n\delta_{k1} \qquad (T11.5)$$

现在，我们必须审查不同 n 值系统的标准(T11.2)。在单形 S_n 的内域，条件$\dot{V}\leqslant 0$ 变得与如下不等式等价：

$$r(\zeta)<\frac{1}{n} \qquad (T11.6)$$

显然，我们找到 $r(\xi_0)=\frac{1}{n}$，它满足方程$\dot{V}(\xi_0)=0$。(ξ_0代表了此超循环的中心固定点)。

对于 2 维系统($n=2$)，条件(T11.6)容易得到验证：

$$\zeta_1=\zeta;\quad \zeta_2=1-\zeta\rightarrow r=2\zeta(1-\zeta)\leqslant\frac{1}{2} \qquad (T11.7)$$

函数 r 代表了在 $\zeta=\frac{1}{2}$处具有极大值的一条抛物线。所以除了在固定点ξ处有$\zeta=\frac{1}{2}$以外，$r(\zeta)<\frac{1}{2}$到处都得到满足。在这种情况下，V 是一个严格的李雅普诺夫函数，$\bar{\xi}_0$是渐进稳定的。

对于$n=3$，情况是非常相似的。不等式$(T11.3)$，即$r<\frac{1}{3}$，除了在固定点$\bar{\xi}_0$处$r=\frac{1}{3}$以外，它在单形S_3的内域中的每一点都保持有效。V又代表了一个严格的李雅普诺夫函数，中心固定点$\bar{\xi}_0$是渐进稳定的。

在4维中，问题就较复杂了。我们认识到，几乎在单形S_4的每一处，条件$(T11.3)$得到了验证：

$$r=(\zeta_1+\zeta_2)(\zeta_3+\zeta_4)=S(1-S),\quad 0\leqslant S\leqslant 1 \qquad (T11.8)$$

在其内域，我们发现$0\leqslant r\leqslant\frac{1}{4}$，当且仅当$S=\frac{1}{2}$时，$r=\frac{1}{4}$。方程$S=\frac{1}{2}$决定了平面$\zeta_1+\zeta_2=\frac{1}{2}$（参见图1-31$a$和图1-34$b$）。$V$显然只是在较弱意义上的一个李雅普诺夫函数。这个结果提示，至少中心固定点是稳定的。为了证明渐进稳定性，我们引入新变量x、y、z:

$$x=-2(\zeta_2+\zeta_3)+1 \qquad \dot{x}=-(1+z)(y-xy)$$
$$y=2(\zeta_1+\zeta_2)-1 \qquad \dot{y}=(1-z)(x-yz)$$
$$z=2(\zeta_1+\zeta_3)-1 \qquad \dot{z}=z^3-z+x^2-y^2$$

这使得坐标系统的原点移到单形S_4的中心，并使得坐标轴分别通过边$\overline{23}$、$\overline{34}$和$\overline{13}$的中点（参见图1-31a）。第4个变量$\zeta_4=1-\zeta_1-\zeta_2-\zeta_3$则被删掉了。从而$z$轴垂直地指向临界平面$\zeta_1+\zeta_3=\frac{1}{2}$，此平面是由$x$和$y$这两个变量张成的。在此平面中，动力学系统简化为$\dot{x}=-y$、$\dot{y}=x$以及$\dot{z}=x^2-y^2$。

z的时间导数只有分别沿着$x=\pm y$或者$\zeta_2=\zeta_4$和$\zeta_1=\zeta_3$这两条线才会消失。因而，在此临界平面中，除了固定点$\bar{\xi}_0$以外，不存在轨道，此系统瞬间就通过了。沿着任一给定的轨道，除了当通过临界平面$\zeta_1+\zeta_3=\frac{1}{2}$的瞬间是例外，几乎在每一时刻条件$\dot{V}[\zeta(t)]<0$都得到满足，沿所有的轨道，$V[\zeta(t)]$都是随时间$t$而单调地减少的。$V$是一个严格的李雅普诺夫函数，从而固定点$\bar{\xi}_0$是渐进稳定的。

在更高的维数$n\geqslant 5$，$V(\zeta)$不再是一个李雅普诺夫函数，从而此方法无法对中心固定点的稳定性作出预言。

在基本超循环中，单形的隅角是鞍点：沿着边 $\overline{h_i}(\delta x_i>0, h=i-1+n\delta_{i1})$，隅角$(i)$关于涨落是稳定的，但沿着边 $\overline{ij}(j=i+1-n\delta_{in})$ 则是不稳定的。在每一个完全的动力学系统的边界上，我们从而发现了一个封闭环 $\overline{12}\ \overline{23}\ \overline{34}\ldots\overline{n1}$，此系统沿该环具有明确的旋转意义。这个循环不是一条单一的轨道。为允许此系统进入下一个纯净态，在每一个隅角要求一种特殊的涨落。存在这个环相当于总系统的循环对称。每一个隅角上的不对称，反映了在我们的模型中假定的生物聚合物的合成和降解的不可逆性。

在动力学系统$3A$中，物理上允许的变量范围是由两条顺序相连的流动边 $\overline{ij}$ 和 $\overline{jk}$ $(j=i+1-n\delta_{in}$ 和$k=j+1-n\delta_{jn})$ 以及一条固定点边 $\overline{ik}$ $(k\neq i-1+\delta_{i1})$ 所包围的。图1-31b 中显示了这个系统的轨道。它们从固定点边的某个点出发，终止于隅角 k，从而它代表了此系统唯一的稳定吸引子。因而物种 I_k 代表了一个超循环的这种片段的生存者或残留者。

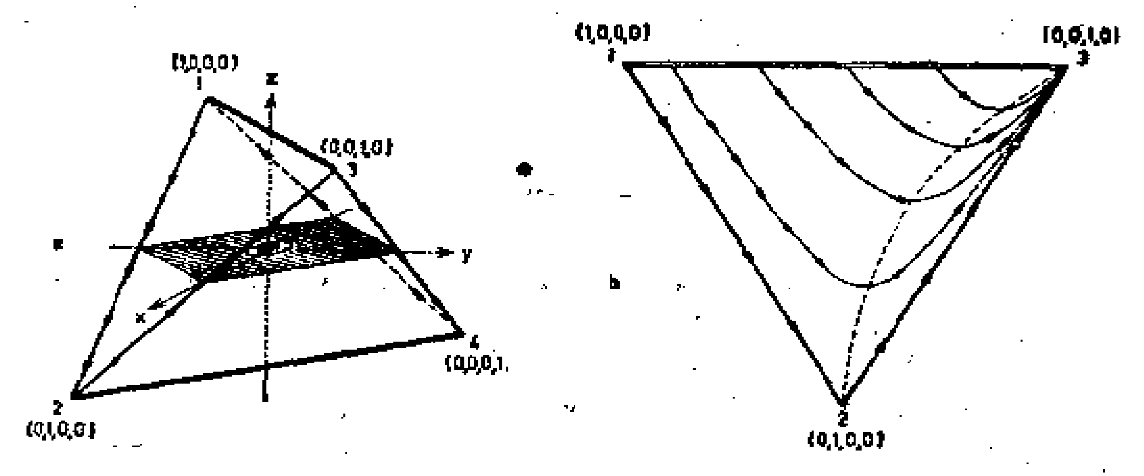


图 1-31　维数$n=4$的基本超循环的动力学拓扑　单形上的这个动力学系统由内域IS_4上的系统4和在等边三角形(S_3)上的4个等价的$3A$型系统构成，其中的每一个$3A$型系统都是由两条流动边$2A$和一条固定点边$2B$包围成的。a)内域上的系统恰当地用变量x、y和z来描述（见表1-11）。封闭同心轨道流形属于x、y平面（图中影线区域）上的线性化系统的中心。b）动力学系统$3A$：所有的轨道都始于固定点边$(\overline{13})$上的某一点，并终止于同一隅角(3)。在虚线连起来的各点上轨道都与固定点边$(\overline{13})$平行。

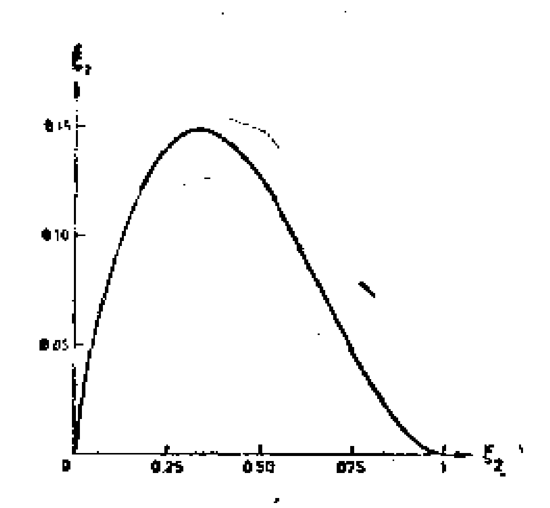


图 1-32 "流动边" 2A 切向量 $\dot{\zeta}_2=\zeta_2(1-\zeta_2)^2$在整个物理学上允许的区域内($0<\zeta_2<1$) 都为正，而在两端则消失。两端代表了此系统的两个固定点：$\bar{\xi}_1=(1,0)$和 $\bar{\xi}_2=(0, 1)$。$\dot{\zeta}_2>0$ 意味着，ζ_2 在下一个时间微分间隔中是增加的。因此，2A 上只有一个轨道从 $\bar{\xi}_1$ 到 $\bar{\xi}_2$，即从隅角 1($\zeta_2=0$)接近隅角 2($\zeta_2=1$)。动力学系统沿着这条边"流动"。注意，在固定点 $\bar{\xi}_2$($\zeta_2=1$)处 $d\dot{\xi}_2/d\zeta_2$ 消失了，导致线性化系统的一个本征值 $\omega=0$。因而固定点分析不能预见这个点的稳定性。

对基本超循环的边界的考察，可以推广到更高维的系统。从所获的结果，我们可以预期破缺的超循环的长期发展。在超循环的一种物种被某一外部事件所消灭后，余下的动力学系统是不稳定的，并在充分长时间后趋近一个纯态。在所有情况下，被选择的某个物种正好在此超循环破缺之前出现。换言之，当 i 是 $j(j=i+1-\delta_{in})$ 的前体时，物种 I_i 将保持超循环的最后残留者，而这个超循环由于其成分I_j灭绝而受到破坏。根据已知的催化链的性质来看，这种行为并不是意外的。

2.4.2 数值积分

维数 $n\leqslant12$ 的基本超循环的微分方程组都已用标准数值技术进行了积分。在以前的论文中已给出了相应的解曲线 $x(t)$，在此就不重复了，

因为我们在此关心的是这个问题的另一个方面。我们现在的目的是研究单形S_n内域中的稳定吸引子，它们保证了组分的协同行为。为此目标，我们直接考察轨道流形。

在原先的动力学系统中去掉显时间相关，就获得了轨道的微分方程：

$$\frac{dx_2}{dx_1}=\frac{A_2}{A_1}=f_2(x_1, x_2, \ldots, x_n)$$

$$\frac{dx_3}{dx_1}=\frac{A_3}{A_1}=f_3(x_1, x_2, \ldots, x_n)$$

$$\vdots$$

$$\frac{dx_n}{dx_1}=\frac{A_n}{A_1}=f_n(x_1, x_2, \ldots, x_n) \tag{70}$$

对这个新的$(n-1)$维动力学系统积分，得到作为解曲线的轨道：

$$x_2=g_2(x_1, x_2, \ldots, x_n)$$

$$x_3=g_3(x_1, x_2, \ldots, x_n)$$

$$\vdots$$

$$x_n=g_n(x_1, x_2, \ldots, x_n) \tag{71}$$

因此，此轨迹是n维浓度空间的一条曲线。为图示出来，我们将把这些曲线投影在由选定坐标x_k与x_1所生成的平面上。低维($n=2$、3 和 4)超循环的轨道反映了这些动力学系统的已知性质。$n=2$的情况是相当平凡的，只有两条收敛于中心稳定焦点的轨道(图1-30)。3 维超循环($n=3$)的轨道是螺线，它迅速趋近中心固定点(图1-33)。这种轨道对应于解曲线$x(t)$的强阻尼振荡。4 元超循环值得深入考虑。此轨道又是螺旋地到达单形中心(图 1-34a、b)。与3 维的例子比较可知，中心力比旋转分量要弱得多。相应地，向中心固定点的收敛是极慢的。轨道在(x_1, x_3)平面的投影很好地说明了前面推导出的结果，即在平面$x_1+x_3=\frac{1}{2}$和$x_2+x_4=\frac{1}{2}$上不存在轨道。的确，正如人们能从图 1-34b所见到的，轨道基本上在一个类似鞍面的弯曲面上。

对于维数$n\geqslant 5$的基本超循环，中心固定点代表了一个不稳定鞍点。

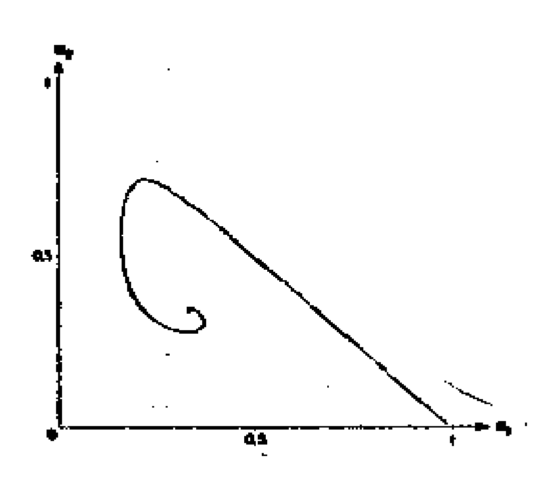
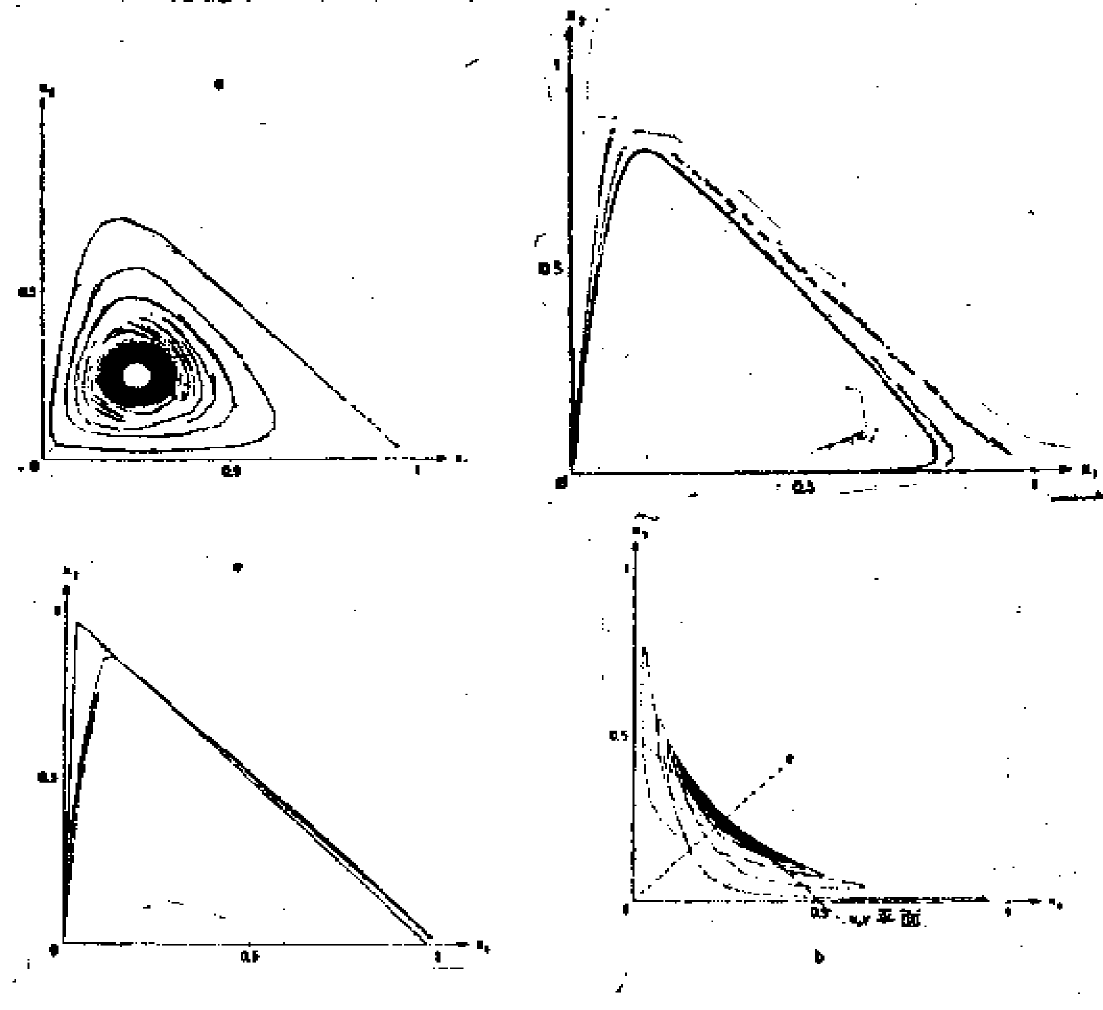


图 1-33 维数 $n=3$的基本超循环动力学系统(3)的轨道，图示为在平面 $(x_1,\ x_2)$ 上的投影。初始条件：$x_1(0)=0.98,\ x_2(0)=x_3(0)=0.01$



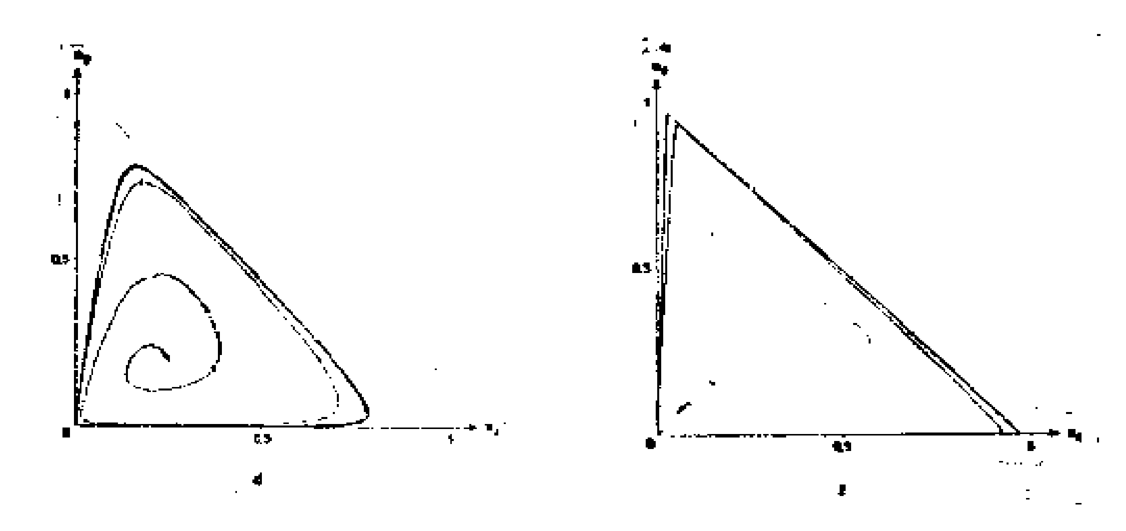


图 1-34 维数分别为 $n=4$、5 和 12 的基本超循环的动力学系统(4、5和12)的轨道 (1)$n=4$，初始条件：$x_1(0)=0.97$，$x_2(0)=x_3(0)=x_4(0)=0.01$；两种投影图：a)在平面$(x_1,x_2)$上的轨道投影，此轨道以强衰减振荡螺旋地进入中心面定点；b)在平面(x_1, x_4)上轨道的投影，中心流形平面(图1-31的x、y平面)与x_2、x_3平面沿直线$x_1+x_3=1/2$相交，垂直于它我们看到 Z 轴：注意，轨迹只与这个平面(x, y)的单点相交，所以不会长时间停留在那里(表 1-11)。(2)$n=5$，投影在平面(x_1, x_2)上：c) 初始条件：$x_1(0)=0.9996$，$x_2(0)=x_3(0)=x_4(0)=x_5(0)=0.0001$：d)初始条件：$x_1(0)=0.2004$，$x_2(0)=x_3(0)=x_4(0)=x_5(0)=0.1999$。注意，对于两组初始条件，此动力学系统都趋近相同的极限环。(3)$n=12$，投影在平面(x_1,x_2)上；e)初始条件：$x_1(0)=0.9989$，$x_2(0)=\ldots=x_{12}(0)=0.0001$；f)初始条件：$x_1(0)=0.0848$，$x_2(0)=\ldots=x_{12}(0)=0.0832$。两个极限环又是相同的，并非常接近于环$\overline{1\ 2}$，$\overline{2\ 3}$，$\cdots$，$\overline{11\ 12}$，$\overline{12\ 1}$。

在边界中没有收点，从而人们期待有一个稳定封闭轨道。对于在单形的内域是否存在这种吸引子，目前的分析技术尚无法给予证明。所以，我们不得不依赖数值结果。

对于一个极限环或封闭轨道，数值积分的确提供了有力的证据。从非常接近单形的中心、某个面、某条边或某个隅角的各种点出发，在充分长的时间以后总是到达相同的极限环。图 1-34 c—f，分别对于维数 $n=5$和 $n=12$ 的基本超循环显示了两条典型轨迹。正如我们可以从这两种图的对比中看到，随着 n 的增加，极限环更为趋近前面提到的圈$\overline{1\ 2}$，

$\overline{23}$，…$\overline{111}$。结果是，单种浓度中的振荡变得越来越类似于矩形脉冲。

运用数值技术，也使我们可以不要如下假设：$k_1=k_2=\ldots=k_n$。对于维数$n=4$和$n=5$的动力学系统，已对任意k值作了计算。但没有观察到解曲线的一般本性有任何变化。典型的例子显示在图1-36中。在两个系统中，单种浓度都是振荡的。对于$n=4$，浓度波是衰减的，且动力学系统趋近中心固定点。它的坐标由如下方程所确定：

$$\bar{x}_0:\overline{x_i^0}=\frac{k_j^{-1}}{\sum_{l=1}^{n}k_l^{-1}}c_0;\quad j=i+1-n\delta_{in} \tag{72}$$

具有不相等速率常数的5元超循环显示了相同种类的无阻尼浓度脉冲，在具有相等k值的系统中已观察到了这种情况。对于所有的亚单元，脉冲的大小不都是相同的。时间平均浓度[由(67)式定义]满足上述(72)式，它决定着(不稳定的)中心固定点的位置。相应地，在速率常数相对小的步骤之前的那些物种的脉冲是宽的，而在相对快的反应步骤之前的那些物种的脉冲，其宽度和高度都较小。从而，此系统以优化的总生产速率调节它的组分浓度。

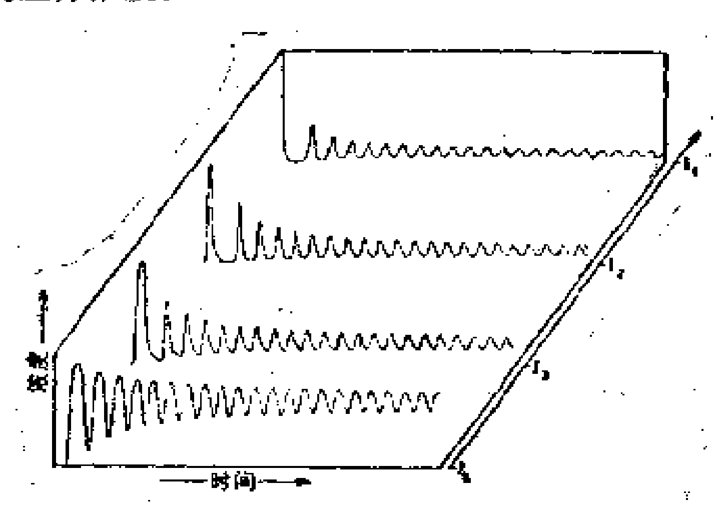


图1-35 一个基本超循环动力学系统的解曲线 它的维数$n=4$，速率常数不相等[$k_1=0.25$，$k_2=1.75$，$k_3=1.25$，$k_4=0.75$；初始条件：$x_1(0)=0.9997$，$x_2(0)=x_3(0)=x_4(0)=0.0001$；满浓度标度=1个浓度单位，满时间标度=1000个时间单位]。注意，I_1(在最快步骤前面的组分)的浓度是最小的，而I_4(在最慢步骤前面的组分)的浓度则是最大的。

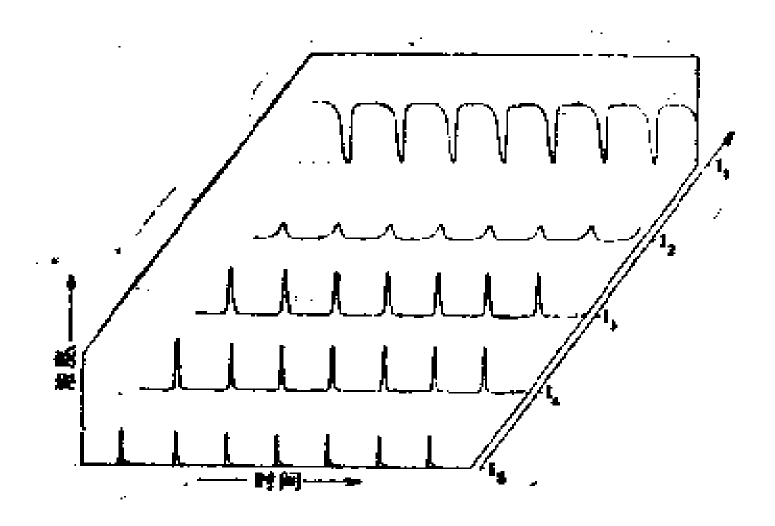


图 1-36　一个基本超循环动力学系统的解曲线　它的维数 $n=5$，速率常数不相等［$k_1=25/13, k_2=1/13, k_3=19/13, k_4=1, k_5=7/13$；初始条件：$x_1(0)=0.9996$，$x_2(0)=x_3(0)=x_4(0)=x_5(0)=0.0001$；满浓度标度＝1个浓度单位，满时间标度＝1200个时间单位］。注意，I_5（最快步骤前的组分）的浓度是最小的，而 I_1（最慢步骤前的组分）的浓度是最大的。

较高维（$n\geqslant5$）的超循环不以恒定的稳定浓度的稳定状态存在，而在中心的一个不稳定固定点周围显示出象波一样的振荡。然而，因为各组分的浓度受到整个系统的动力学所控制，并且群体变量不消失，所以它们表现出协同行为。

对应于基本超循环的动力学系统，在单形内域有且仅有一个吸引子，其注域扩展到所有化合物浓度为正的（非零）整个区域。在低维（$n\leqslant4$）时，此吸引子是一个渐进稳定固定点，即对于 $n=2$ 是一焦点，对于 $n=3$ 和 $n=4$ 是一螺旋收点。在更高维（$n\geqslant5$）系统中，数值积分对于一个稳定极限环的存在提供了有力证据。因此，所有的基本超循环都是由它们组分的协同行为作为特征。

这种类型的超循环，由于它们的动力学特征，所以对于自

组织(耗散结构，即处于叠置输运的情况)还有许多尚未探索过的潜在重要性。它们也许还在神经网络的自组织中起着重要的作用。

2.5 超循环与翻译

2.5.1 理想的边界条件和一般简化

一组合适的边界条件能够在一个流动反应器中得到实现。所有低分子化合物的浓度($m_i, i=1,2,\cdots,\lambda$)，借助于控流装置得到缓冲，它们同时又向此系统提供能量。浓度变量x_i涉及反应器中合成大分子的种类，而所有其他的"标准反应混合物"的化合物，则不在此微分方程中明显地反映出来，而隐含在(30)式的有效速率常数中。

由于技术上的困难，从一些推论中认识到，要详尽理解反应机制中的所有基本步骤是不可能的。我们宁可运用简化的反应模式，从而导致了合适的"总"动力学。在化学动力学中，这种策略是一种普通程序。水溶液中的酸碱反应就是例子；它们一般用唯象方程来描述，不解释其中的单个质子跃迁，但又正好反映了所考虑分子的质子化作用中的状态变化。

对于模板指导的聚合作用和翻译的机制，速率方程的变量只涉及到作为仅有变量的总的大分子群体数。因此，并不详细考虑链的引发和传播步骤。从实验上可以证明，这些近似是正当的。实际上，我们在运用的这种"总"动力学业已确立(参见第三部分)。

2.5.2 动力学方程

图1-37 所示意的催化超循环，由两组大分子构成：n 个核苷酸和 n 个多肽。聚核苷酸(I_i)的复制，是由多肽(E_i)催化的，多肽反过来又是前者的翻译产物。超循环连接是由两类动力学相互关系建立起来的。

1．每一个聚核苷酸I_i被专一地翻译成为一个多肽E_i。翻译的可能性显然需要一种合适的机制，这种机制至少由一些翻译产物E_i所组成，同时还运用着一定的遗传密码。

2．聚核苷酸和多肽生成的特殊复合物，它们在聚核苷酸拷贝的合成中又具有催化活性。多肽可能是某种具有聚合酶活性的普通蛋白质的特殊复制酶或特殊辅助因子。总起来说，这些原始蛋白质至少提供了两种功能：特定的复制和翻译。第三部分中表明如何把这种系统形象化。

在I_i和E_i之间的耦合必须具备如下形式：它允许反馈环闭合(图 1-37)。用数学术语来说即是，假定在酶E_j和聚核苷酸I_i(在此 $j=i+1-n\delta_{in}$) 之间形成特定的复合物，从而引入了循环对称。

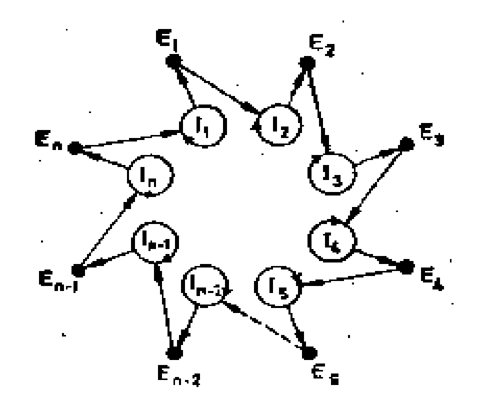


图 1-37 带翻译的超循环模式图 维数为 $2\times n$，即n个聚核苷酸和 n 个多肽。

聚核苷酸合成的动力学遵循米凯利斯-门顿型反应模式，尽管我们并没有假定复合物浓度小得可忽略不计。

$$I_i+E_i \underset{}{\overset{k_1}{\rightleftharpoons}} I_iE_i;$$

$$I_iE_i+\sum\nu_\lambda^i M_\lambda^i \xrightarrow{f_i} I_i+I_iE_i \tag{73}$$

M_λ^i 和 ν_λ^i 表示4个核苷三磷酸及其计量系数，其中分别有 $\lambda=1, 2, 3, 4$。现在我们引入 z_i 表示复合体 I_iE_i 的浓度，用 x_i^0、y_i^0 或 x_i、y_i 表示多肽 (E_i) 和聚核苷酸 (I_i) 的总浓度或自由浓度。质量守恒要求：

$$x_i^0=x_i+z_i \quad 和 \quad y_i^0=y_i+z_i \tag{74}$$

对于复合体的快速平衡，浓度 z_i 与总浓度 x_i^0 和 y_i^0 相联系：

$$z_i=\frac{y_i^0+x_i^0+K_i}{2}\left[1-\sqrt{1-\frac{4y_i^0x_i^0}{(y_i^0+x_i^0+K_i)^2}}\right] \tag{75}$$

假定多肽的合成是非专一的，即多肽 I_i 的翻译是借助于某个共同的"翻译器"而出现的：

$$I_i+\sum_\lambda \nu_\lambda^E M_\lambda^E \xrightarrow{k_1} I_i+E_i \tag{76}$$

M_λ^E 和 ν_λ^E 分别表示激活氨基酸及其计量系数。通过分别控制两种生物聚合物（I 和 E）的总浓度，可适当地引入选择约束条件。模仿恒组织约束条件，我们使两种浓度和都保持为恒量：

$$\sum_k y_k^0=C_I^0; \quad \sum_k x_k^0=C_E^0 \tag{77}$$

在所有这些条件下，$2n$ 个耦合微分方程构成的动力学系统写作：

$$\dot{y}_i^0=f_iz_i-\frac{y_i^0}{C_I^0}\cdot\sum_{R=1}^n f_Rz_R$$

$$\dot{x}_i^0=R_iy_i-\frac{x_i^0}{C_E^0}\cdot\sum_{R=1}^n R_Ry_R$$

$$i=1, 2, \ldots, n \tag{78}$$

为达到我们的目的，讨论两种极限情况就足够了：

1．复合体I_iE_j的浓度Z_0变得与聚核苷酸浓度同多肽浓度的乘积成正比，这时浓度应充分低：

$$y_i^0, x_j^0 \ll K_i; z_i \sim \frac{1}{k_i} y_i^0 x_j^0; z_i \ll y_j^0, x_j^0 \tag{79}$$

如果我们进一步假定与二级复制的速度相比，一级翻译过程的速度是快的，这个假定至少对低浓度聚核苷酸来说是正的。多肽浓度将来取定常值，它可被包括在速率参数中；于是，聚核苷酸的生成便是一个典型的n维基本超循环，可由一个微分方程组来描述。

2．在高浓度时，z_i变得与两个变量中的较小一个相等；①

$z_i^0 = inf\ (y_i^0,\ x_j^0)$

相应地，我们趋近了两种可能的极限情形：

$$K_i \ll y_i^0 \ll x_j^0: \quad z_i \sim y_i^0 \tag{81}$$

$$K_i \ll x_j^0 \ll y_i^0: \quad z_i \sim x_j^0 \tag{82}$$

其中的第1种情况，聚核苷酸的行为象独立的竞争者，而由于$y_i = y_i^0 - z_i = 0$，多肽保持稳定。在自然条件下，象"恒总浓度"那样的约束条件在此不合适，至少对于小的y值不合适，结果是，聚核苷酸的生长将导致浓度比$y:x$的倒换，从而趋近条件(82)。于是，如下近似

$$z_i \approx x_j^0 \ \text{和}\ y_i \approx y_i^0 \tag{83}$$

成为有效，因而导致了一个$2n$元的催化循环，但不是一个超循环。从而在饱和条件下，各组分处于高浓度时，超循环失掉了典型的非线性生长速率行为。作为一个统一的系统，它模拟了一个简单催化循环的性质，催化循环等价于自催化剂或自复制单元。

① inf＝下确界，这是一个数学术语，代表一个集合中的最小成员。

2.5.3 数值解

对于详细考察聚核苷酸和多肽之间形成复合体所导出的催化超循环的微分方程，难以用分析方法来研究，因为其中涉及到无理表达式。在这些情况下，数值积分虽是费事的，但是只有它才能描述关于这些动力学系统信息的仅有的信息源。为了例示聚核苷酸-多肽超循环的动力学，我们将通过计算机给出解曲线和轨道。

与基本超循环比较，聚核苷酸-多肽系统包含了一类新的参数，亦即复合体的结合常数K_i。正如从处于低浓度极限和高浓度极限的动力学行为的差异所预期的，平衡常数表现了对系统的动力学性质的某种支配性的影响。为了进行系统的考察，我们减少独立参数的个数。所作的这些假设与基本超循环中所用的假设是基本相同的：即对于聚核苷酸复制的全部速率常数，有$f_1=f_2=\ldots=f_n=f$；对于它们翻译成多肽，有$k_1=k_2=\ldots=k_n=k$；以及所有的结合常数假定相等，即$K_1=K_2=\ldots=K_n=K$。然后，我们研究在一组恒定起始浓度下，K对处于固定的f值和k值的动力学性质的影响。对于维数$n\leqslant4$的超循环，在充分长时间后，解曲线趋近一个稳定的定态。单个浓度可能显示出阻尼的振荡。这些系统的动力学，本质上与具有更高的n值和小的平衡常数的超循环是相同的。

较高维超循环的动力学更为复杂。系统的长期行为随平衡常数K值的增加而变化。在一定的临界值(K_{cr})以下，系统向稳定的定态收敛，而对于较大的K值($K>K_{cr}$)则获得极限环。根据解曲线和轨道的形状，我们区分4种不同的情况，按照平衡常数K值的增加来排列：

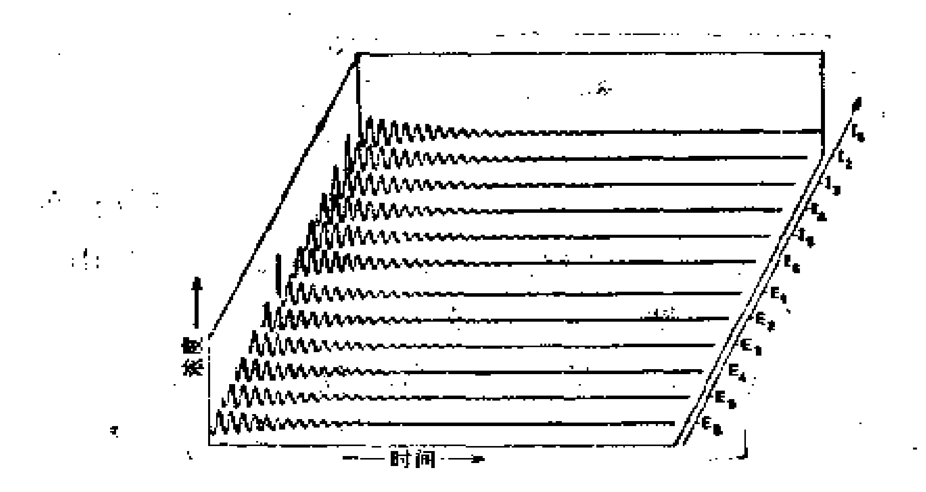


图 1-38 带翻译超循环的动力学系统的解曲线 维数2×6，$K=0.25$；初始条件：$y_1(0)=5.0, y_2(0)=\dots=y_6(0)=0.5$；$x_1(0)=\dots=x_6(0)=1.0$；满浓度标度=5 个浓度单位，满时间标度=1000 个时间单位。对于雷 普 夫 分支，平衡常数 K 值低于临界值，因此观察到一种阻尼振荡。

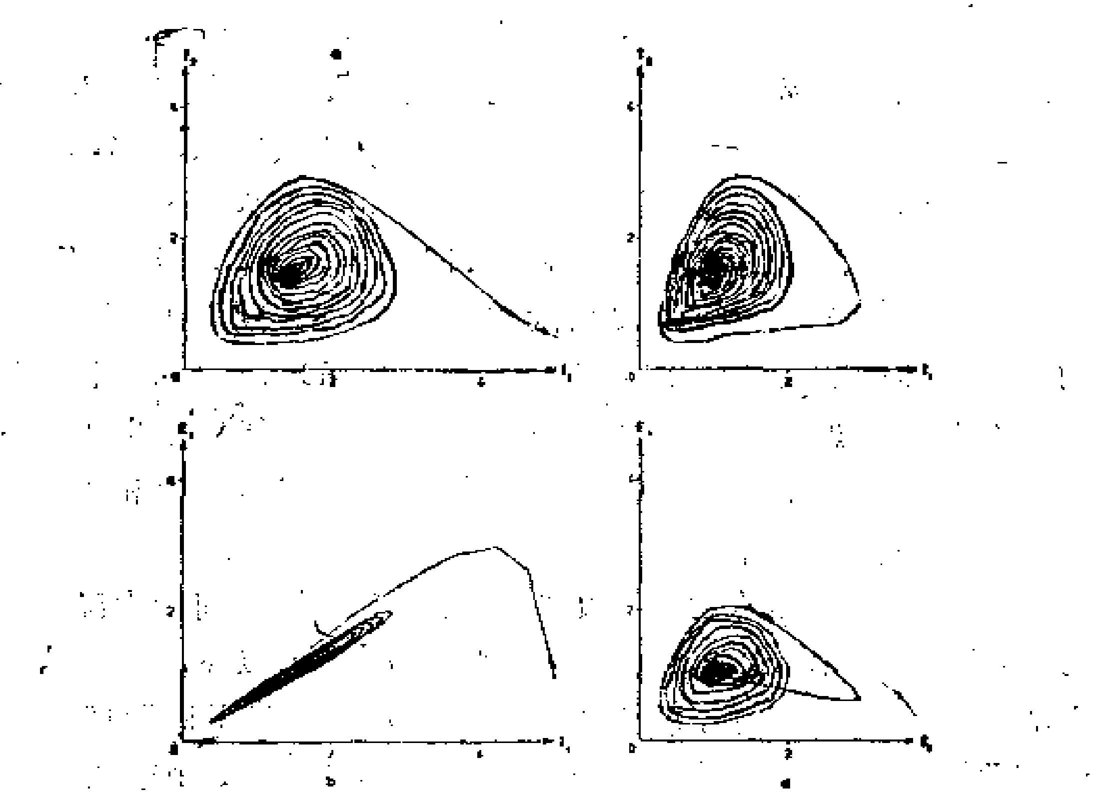


图 1-39 带翻译超循环的动力学系统的轨道 维数2×5，$k=1.0$；初 始 条 件：

$y_1(0)=5.0, y_2(0)=y_3(0)=y_4(0)=y_5(0)=0.5; x_1(0)=x_2(0)=x_3(0)=x_4(0)=x_5(0)=1.0$。$a$)轨道在平面($y_1$, y_2)上的投影表示了聚核苷酸 I_1 和 I_2 的浓度。b) 平面(y_1, x_1)上的投影表示了聚核苷酸 I_1 及其翻译产物酶 E_1 的浓度。注意，E_1 的浓度大致与 I_1 的浓度成比例，所以，简化带翻译的超循环条件是能满足一种良好的近似的。c)在平面(x_1, y_2) 上的投影显示了多肽 E_1 和聚核苷酸 I_2 的浓度，I_2 的形成是前面的 E_1 催化的。d) 在平面(x_1, x_2)上的投影，显示了多肽 E_1 和 E_2 的浓度。注意，K 又低于霍普夫分支的临界值，此轨道向中心固定点收敛。

1．在K值较小时，动力学行为与低维数超循环在定性上相同。解曲线显示了强阻尼振荡(图1-38)，轨道螺旋式地迅速落入代表稳定的定态中心(图1-39)。

2．原则上，我们发现了与情况(1)相同的一般类型的动力学行为。不过振荡阻尼很小，向定态趋近得极慢(图1-40a、b)。这种情形与 (1) 大不一样，因为在简正模分析中不出现衰减项，但要考虑非线性贡献。这一事实唯象地揭示了它自己的初始振幅(几乎)是恒定的。在平衡常数K略小于临界值 K_{cr}时，即 $K=K_{cr}-\delta K$，就出现这种情形。

3．在K值略大于临界平衡常数($K=K_{cr}+\delta K$)时，我们观察到一种有趣的现象。此动力学系统开始的行为很象情况(2)。单个浓度以相对小的幅度振荡。与情况(2)比较，在初始期间振幅稍有增加。然而，过了这个正弦振荡阶段后，浓度波陡然地改变其形状和频率(图1-40c、d)，于是变得很象矩形脉冲，这种脉冲我们已在高维的基本超循环中遇到过。最后，此动力学系统趋近一个极限环。

4．在K值较大时，单个浓度的振幅增加，动力学系统稳步地趋近极限环(图1-40e、f)。我们在此观察到的具有连续变化参数的动力学行为的这种变化，在文献中被称作"霍普夫分支"。这种向着长期解收敛中的特殊减速，通常叫做霍普夫分支的"临

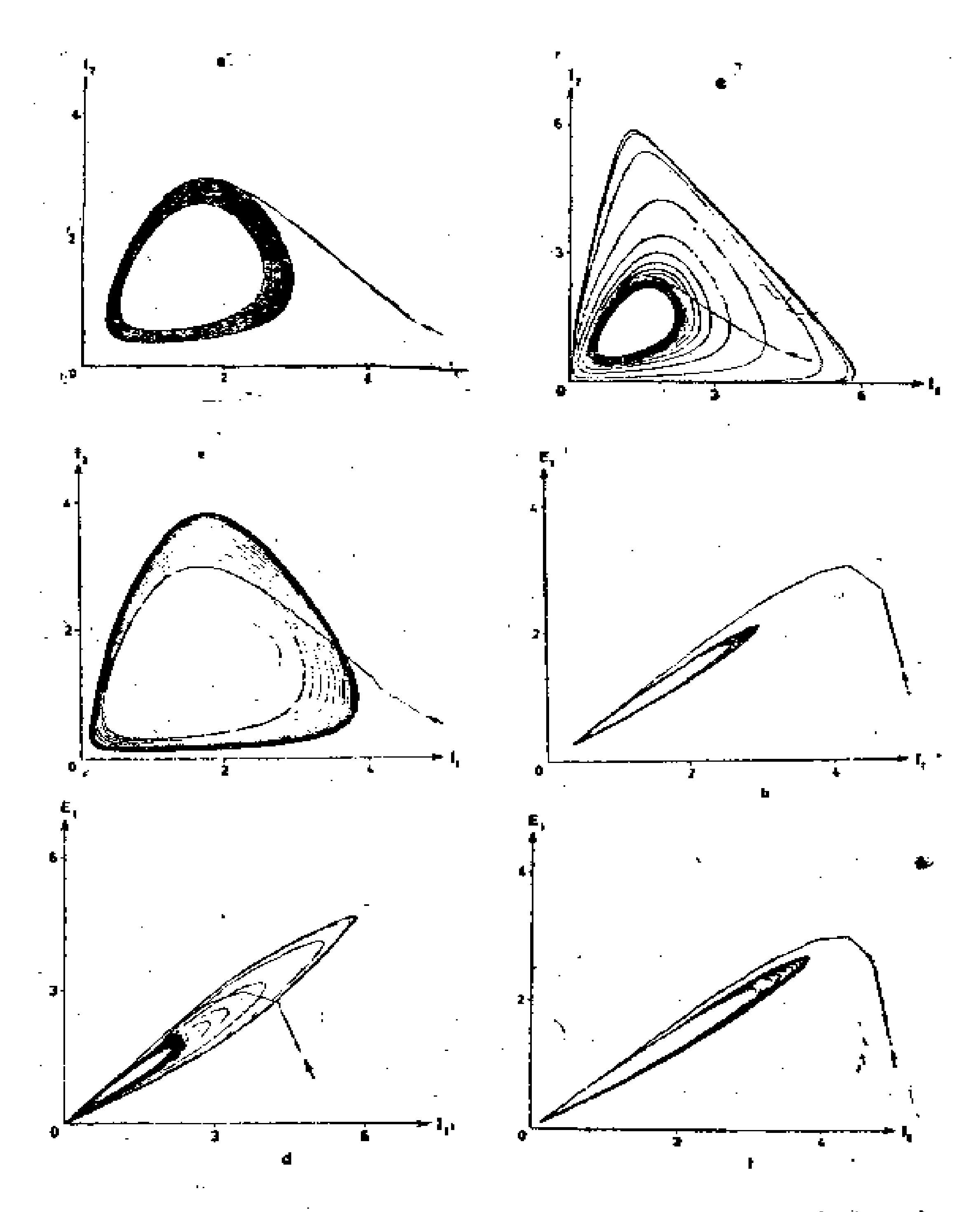


图 1-40　带翻译超循环的动力学系统的轨道　a)和 b)的维数 2×5, $K=1.1$; 初始条件: $y_1(0)=5.0$, $y_2(0)=y_3(0)=y_4(0)=y_5(0)=0.5$, $x_1(0)=x_2(0)=x_3(0)=x_4(0)=x_5(0)=1.0$; 此平面上的投影分别代表了 I_1、I_2 和 I_3

E_1 的浓度；平衡常数K的值稍低于霍普夫分支，我们观察到，向稳定的中心固定点收敛得非常慢。c)和 d)维数 2×6，$K=0.2784$；初始条件：$y_1(0)=5.0, y_2(0)=\ldots=y_6(0)=0.5, x_1(0)=\ldots=x_6(0)=1.0$；在浓度为 I_1、I_2和 I_1、E_1 的平面上分别显示了它们的投影；平衡常数 K 的值稍大于霍普夫分支，我们观察到，系统最终收敛到稳定极限之前还有一个亚稳极限环。e)和 f)的维数 2×5，$K=1.2$；初始条件：$y_1(0)=5.0$，$y_2(0)=y_3(0)=y_4(0)=y_5(0)=0.5$，$x_1(0)=x_2(0)=x_3(0)=x_4(0)=x_5(0)=1.0$；在浓度为 I_1、I_2 和 I_1、E_1 的平面上分别显示了它们的投影；平衡常数 K 的值在霍普夫分支之上，此系统稳步地向稳定极限环收敛。注意，在 E_1 和 I_1 之间的比例，在所有 3 种情况中（b、d和f）都颇合理地得到了满足。

界迟缓"，我们在情况(2)和情况(3)中发现过，也已用来描述了其他一些动力学系统。在超循环情况中，临界K 值附近的"迟缓"对于大的n值变得更为显著。在 5 元循环中（$n=5$），某种对应于(3)的情形是难以检测的。另一方面，$n=10$ 的催化超循环比

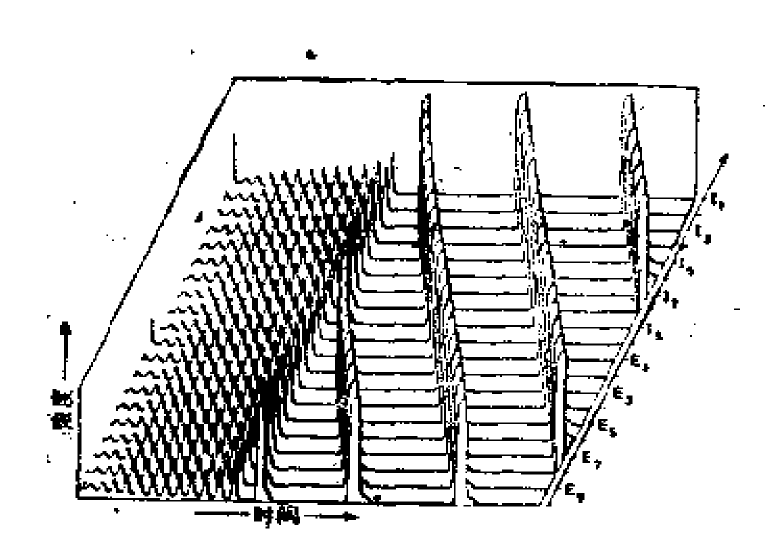


图 1-41 **带翻译的超循环动力学系统的解曲线** 维数 2×10，$K=0.026$；初始条件：$y_1(0)=5.0, y_2(0)=\ldots=y_{10}(0)=0.5, x_1(0)=\ldots=x_{10}(0)=1.0$；满浓度标度=10 个浓度单位，满时间标度=1000 个时间单位。选择的平衡常数值稍在霍普夫分支的临界值之上。我们观察到一个亚稳振荡态，它带着特殊的浓度波突然转变到最后的极限环。

起6元系统(图1-41)，则表现出长得多的起始期，这正如情况(3)中的情形。起始的正弦振荡阶段，类似于某种亚稳振荡态。它在向终态极限环过渡时，随着n值增加变得尖锐起来，这在10元超循环时相当明显。

迄今研究过的所有聚核苷酸-多肽超循环，在物理上允许的浓度范围的内域中都有一个吸引子。它们以其组分的协同行为作为特征。依据总浓度(C_0^N和C_0^P)的积、缔合常数(K)的值以及超循环的大小，我们观察到稳定的固定点或者极限环。于是小的K值与高的浓度互补，反之亦然。在低浓度极限和高浓度极限的长期行为，可由数值积分获得，它与上一节在分析基础上作出的预见完全一致。对涉及到多肽合成的准稳定态性质的基本简化，可以由观察在E_1、I_1平面上的轨道的投影而直接核实。对于此定态近似，我们期望发现直线。正如我们能够从图1-39b和图1-40b、d、f看到的，两种浓度的比例大致上得到满足，简化处理显得是很正当的。实际上，对这种复杂反应机制进行数值分析的目的是证明，复杂的超循环与基本的超循环就其自组织性质而论是等价的。因此，由基本系统所获得的结论，是与所有各种结构更复杂的现实超循环相关联的(参见第三部分)。

2.6 超循环网络

2.6.1 内平衡和超循环之间的竞争

在2.2节中引入的内平衡概念，看来对于直接分析更复杂的组织网络是十分有用的，因为它允许减少独立变量的个数。

首先，让我们考察一下基本超循环中的平衡过程。为此，我

们计算单个浓度 $X_i(t)$ 的时间平均值[参见(67)式]，并把它们与相应的解曲线 $X_i(t)$（图1-42）作比较。无论终态是定态惰性的还是定态振荡的，经过几个循环后，时间平均值 $\bar{X}_i(t)$ 实际上成为恒量。因此，已建立内平衡的假定看来对于超循环是一个很恰当的近似。然而，我们应在几种情况下对此进行检验。

运用内平衡概念，我们能够为整个超循环的净生长速率导出一个方程：

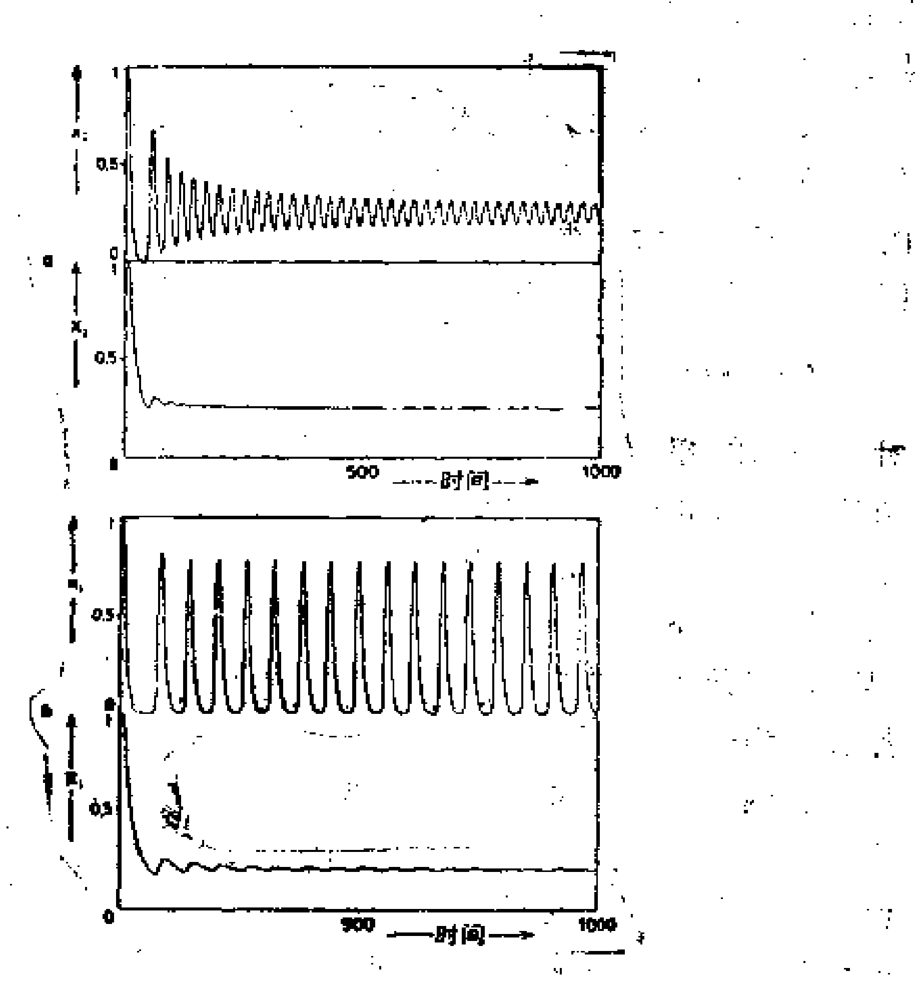


图 1-42 在具有相等的速率常数和时间平均浓度 $X(t)$ 时，维数 $n=4$ 和 $n=5$ 的基本超循环动力学系统的解曲线：(a) $n=4$，以及 (b) $n=5$。注意，经过 n 次振荡以后 $X(t)$ 到达 $\bar{x}$，亦即两个例子中内平衡的建立都是相对快的。

$$\dot{c} = \sum_{i=1}^{n} \dot{x}_i = \sum_{i=1}^{n} \Gamma_i(x) = \sum_{i=1}^{n} k_i x_i x_j$$

$$= \frac{1}{\sum_{i=1}^{n} k_i^{-1}} c^2 \equiv \bar{k} c^2 = \Gamma(c) \qquad (84)$$

$$j = i + 1 - n\delta_{in}$$

因此，超循环是以二次生长速率为特征的，并遵从双曲线生长定律。对于2.2节和2.3节讨论过的非达尔文的"一旦则永存"选择，它们是合适的例子。

按照(84)式$\bar{k}$的表达式，作为整体的超循环的速率常数，将与它的最慢一步的速率常数具有相同的数量级。在无限制生长的条件下，假设超循环在一定的临界时间（t_∞）的生长趋于无限。在完全平衡的系统中，这些不稳定性出现于下式：

$$t_{\infty}^{eq} = [\bar{k}c(t=0)]^{-1} \qquad (85)$$

对于平衡的超循环，按(85)式计算的结果，可与表1-12中由数值积分所获得的远离内平衡(t_{∞}^{neq})系统的值相比。在完全平衡的系统中，不稳定性总是出现得早一些，即$t_{\infty}^{eq}<t_{\infty}^{neq}$。总体上讲，这些数值差异不甚重要，动力学系统的一般行为和t_∞的相对值都得到正确预见。从而关于内平衡的假定对于大多数非平衡系统都表现为一种良好的近似。

在内平衡假设下，全部超循环之间的选择可以作为单个实体进行一般的研究。当然，由此所获得的动力学系统，与那些由二次生长速率所标征的、描述了独立竞争者的动力学系统是一致的。对非平衡超循环之间的竞争更难以进行考察，因为只可能对此种微分方程组进行数值积分。在其他地方讨论了一个例子，它表明，内平衡假设是一种行之有效的近似。

表1-12 在无限制生长条件下超循环动力学系统中的不稳定性

维数	边界和初始条件			临界时间常数	
n	速率常数 k	初始浓度 $c(0)$	初始分布[a] $x(0)$	处于平衡 t_∞^e	远离平衡 t_∞^n
2	1/2	0.55	(0.5,0.05)	3.64	5.0
3	1/3	0.60	(0.5,0.05,0.05)	5.00	6.8
4	1/4	0.65	(0.5, 0.05, 0.05, 0.05)	6.15	7.3

a. 应用在远离平衡系统的数值积分中的初始浓度分布：$x(0)=[x_1(0), x_2(0), \ldots]$

作为竞争系统的一个例子，我们考虑两个超循环 H_A 和 H_B，它们分别是 n_A 元的和 n_B 元的，并服从恒组织约束条件。

如果存在着内平衡，此系统归结为具有二次生长速率项的两个竞争者。由固定点分析得到的结果使我们想起，当其相对初始浓度 $C_A(0)$ 超过了一个临界极限时，超循环 H_A 将被选择：

$$\lim_{t\to\infty} c_A(t)=c_0 \qquad 若\ c_A(0)>\frac{k_B}{k_A+k_B}c_0 \tag{86}$$

否则 H_B 将在竞争中获胜。

考虑一种更为特殊的情况似乎更可说明问题。我们假定，各个速率常数在给定的超循环中是非常相似的，即 $k_1\sim k_2\sim\ldots\sim k_n=\bar{k}_A$ 和 $k_{n+1}\sim k_{n+2}\sim\ldots\sim k_{n+m}=\bar{k}_B$。于是获得如下的全部超循环速率常数：

$$k_A=\frac{1}{n_A}\bar{k}_A,\quad k_B=\frac{1}{n_B}\bar{k}_B \tag{87}$$

正如我们所见，这些常数与超循环中的成员数成反比，结果是较小的超循环似乎具有一定的选择优势。不过，如果我们假定所有的大分子都以大体上相同的浓度 $(\bar{x})$ 出现，那么较大的循环劣势正好由较大的总浓度 C 加以补偿：

$$c_A(0)=n_A\cdot\bar{x},\ c_B(0)=n_B\cdot\bar{x},\ c_0=(n_A+n_B)\cdot\bar{x}$$

$$\lim_{t\to\infty} c_A(t)=c_0 \quad 若\ \bar{k}_A>\bar{k}_B \tag{88}$$

所以，倘若各个成员的初始浓度以及复制步骤的速率常数相等，那么不同大小或者不同维数 n 的超循环的生存机会大致是相同的。

从两个超循环所获的结果，容易推广到 N 个独立的竞争者。

2.6.2 寄生耦合和催化网络

把某个超循环中的所有活性成员 $I_1, \ldots, I_n$ 联系起来的循环封闭的催化连接，也许包括一些分支点，并促进外部物种 $I_{k+1}, \ldots, n$（它们不是此协同单元的内在部分）的进一步增生。我们把这些外部成员叫做寄生者。为使分析处理成为可能，我们应假定此循环中的内平衡(表1-13)。我们用固定点方法对描述了一个超循环和一个寄生者的两个动力学系统进行了研究。

第 1 个例子是一个超循环和一个没有自复制能力的寄生者(图1-43a)。我们从表1-13可见，在总浓度的一定阈值之上($k_A c_0 > k$)，超循环和寄生者在定态以非零浓度出现。此超循环的平衡浓度随 c_0 的增加而增加，而寄生者的浓度则保持不变。于是，在浓度充分高时，对于此循环的动力学，寄生者将彻底失去其重要性。在总浓度($k_A c_0 < k$)较低时，系统变得不稳定。在假定的内平衡限度以内，寄生者破坏了超循环，并最终代表了此动力学系统仅有的残留者。

第 2 个例子描述了带自复制寄生者的超循环的发展（图1-43b)。这个动力学系统以依赖于速率常数 k 和 k_A 的相对值的明确选择为特征。对于 $k > k_A$，寄生者破坏超循环，而不等式 $k < k_A$ 则意味着寄生者的死亡。在单个聚核苷酸水平上详细地考察此动力学系统，也许是有趣的。从表1-13，我们得到

$$k = k_x \frac{x_\nu}{c_A} = k_x \frac{k_{\nu+1}^{-1}}{\sum_i k_i^{-1}} \tag{89}$$

在建立了内平衡的条件下，运用前面导出的表达式

$$k_A=(\sum_i k_i^{-1})^{-1}$$

我们发现：

$$k<k_A\longrightarrow\frac{k_x}{k_{\nu+1}}\cdot\frac{1}{\sum\limits_i k_i^{-1}}<\frac{1}{\sum\limits_i k_i^{-1}}\longrightarrow k_x<k_{\nu+1} \quad (90)$$

从而对于从分支点开始的反应步骤，选择的结果完全由两个速率常数之比决定，而无论此超循环中其他的速率常数的值怎么样。

已对图 1-43 所示的此类动力学系统进行了数值积分，以便研究偏离平衡分布的影响。事实上，所有在此得到的结论，也可在远离平衡的系统中得到证实。

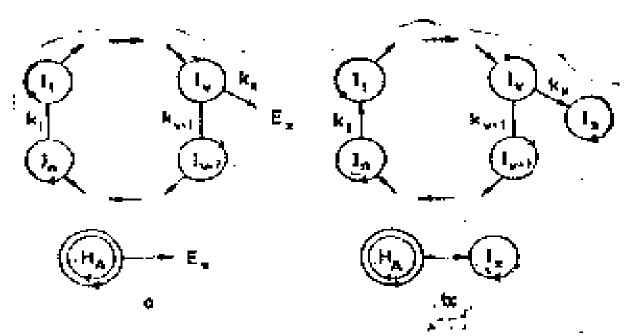


图 1-43　带寄生单元超循环的模式图　*a*）寄生者不是自复制的；*b*）寄生者是自复制的。假定分支出现在组分 ⓥ 处。

表1-13　带寄生偶合超循环的固定点分析

一个带某个寄生单元 I_x 的超循环（图 1-43），由 $n+1$ 个微分方程来描述，其中 n 与孤立超循环中的方程是一致的。第$(n+1)$个微分方程定义了寄生者的浓度，我们有：

$[I_x]=x$，　$[I_i]=x_i$

对于图 1-43*a* 示意的系统有：

$$\dot{x}=k_v x_v-\frac{x}{c_0}\phi \tag{T13.1}$$

对于图1-43b 中的例子有：

$$\dot{x}=k_x x_v x-\frac{x}{c_0}\phi \tag{T13.2}$$

如果在此超循环中已建立了平衡，则 $n+1$ 个微分方程减少为2个。在此引入新的速率常数显得是适宜的：

$$k=k_x\cdot\frac{x_v}{c_A};\qquad c_A=\sum_{j=1}^{n}x_j \tag{T13.3}$$

式中 k_A 由方程式(84)所定义。

(1) 寄生者不是自复制的(参见图1-43a)：

$$\dot{c}_A=k_A c^2_A-\frac{c_A}{c_0}(k_A c^2_A+kc_A);\ c_0=c_A+x \tag{T13.4}$$

$$\dot{x}=k\,c_A-\frac{x}{c_0}(k_A c^2_A+kc_A)$$

这个动力学系统有两个固定点：

$$\bar{x}_1:\ \bar{c}_A=\frac{k_A c_0-k}{k_A},\quad \bar{x}=\frac{k}{k_A} \tag{T13.5}$$

$$\omega^{(1)}=-\frac{(k_A c_0-k)^2}{k_A}$$

$$\bar{x}_2:\ \bar{c}_A=0,\ \bar{x}=c_0;\ \omega^{(2)}=0 \tag{T13.6}$$

如果总浓度不满足临界条件 $c_0=k/k_A$，则 $\bar{x}_1$ 是稳定的。

对 $\bar{x}_2$ 的稳定性进行分析要求详细地考察更高级的项。

对于某个点 $x=c_0-\delta x$，我们发现：

$$\dot{x}=\frac{(\delta x)^2}{c_0}(k-k_A c_0+k_A\delta x) \tag{T13.7}$$

这导致

$\dot{x}>0$　　对于$c_0<k/k_A$

以及

$\dot{x}<0$　　对于$c_0>k/k_A$

因此，固定点 $\bar{x}_2$ 在浓度低于临界值时是稳定的，在浓度高于临界值时是不稳定的。在低浓度时，$\bar{x}_1$ 落在物理上允许的范围之外，$\bar{x}=c_0$ 是唯一的稳定的定态，此超循环被寄生者所破坏。在高浓度时，稳定的固定点 $\bar{x}_1$ 落在单形 S_1 上，这意味着超循环和寄生者是共存的。

(2) 寄生者是自复制的(图 1-43b)：

$$\dot{c}_A = k_A c^2_A - \frac{c_A}{c_0}(k_A c_A^2 + k c_A x), \quad c_0 = c_A + x$$

$$\dot{x} = k c_A x - \frac{x}{c_0}(k_A c^2_A + k c_A x) \tag{T13.8}$$

系统在 S_2 的隅角有两个固定点：

$$\bar{x}_1: \ \bar{c}_A = c_0, \ \bar{x} = 0, \ \omega^{(1)} = (k - k_A)c_0 \tag{T13.9}$$

$$\bar{x}_2: \ \bar{c}_A = 0, \ \bar{x} = c_0, \ \omega^{(2)} = 0 \tag{T13.10}$$

假若 $k_A > k$，则第 1 个固定点是稳定的。对于第 2 个固定点，又需要检测高级项。在 $x = c_0 - \delta x$ 处，我们发现

$$\dot{x} = \frac{k - k_A}{c_0}(\delta x)^2(c_0 - \delta x) \tag{T13.11}$$

现在，这导致

$\dot{x} > 0$　　对于 $k > k_A$

以及

$\dot{x} < 0$　　对于 $k < k_A$

因此，如果不等式 $k > k_A$ 成立，$\bar{x}_2$ 是稳定的。此系统是竞争的，这意味着，除了在速率常数相等 ($k = k_A$) 的特定情形以外，超循环单元与寄生单元是不能共存的。

运用2.3.6节导出的结果，对一元寄生者所获的结果能够推广到任意链。一般说来，所有寄生者的命运与此循环中物种的发展紧密相关：当分支点后面的物种其浓度趋近零时，寄生者总是要死亡的。$k_x = k_{n-1}$，这是一个有趣的特殊情况。I_{n+1} 和 I_x 的微分方程是一致的，因此两个物种的比率总是保持为起始比率而恒定。对 n 个这种类型的动力学系统进行数值积分表明，在这种特殊情形($k_x = k_{n+1}$)中，除了物种 I_x 以外的所有其他寄生成员都将死亡。

链状的寄生者也许在超循环上回叠，由此导向一个具有一个分支点和一个合流的催化网络。通过数值积分，我们发现这类系统是不稳定的：低效率分支即速率常数 k 值较小的分支将死亡，保留下来的是一个单一的简单超循环。

为了将催化耦合项任意分配到一组自复制大分子，我们将

会碰到高度分支的系统或复杂网络，这比规则的超循环要频繁得多。因此，为对超循环形成的可能性作出估计，知道这些系统的进一步发展具有极重要的意义。分析方法通常不适用这种系统，因此我们不得不依赖于数值技术的结果。

对各种催化网络的微分方程进行数值积分获得各种解曲线，由这些解曲线导出了一些普通的结果。正如前面的例子所提示的，这些系统是不稳定的，会解体成为较小的片段。除复杂的动力学结构以外，存在这类结构归因于不同速率常数值的碰巧一致，自复制单元催化网络仅有的可能残留物是独立生长的物种、催化链或催化超循环。因此，由具有均一耦合项的自复制单元构成的任何催化网络都将解体，或是产生出优于其他片段的某个超循环，或是给出不宜协同进化的竞争动力学系统。

2.6.3 超循环之间耦合的等级结构

在原则上，通过向速率方程中引入合适的催化项，可以使超循环耦合起来产生出组织程度更高的系统。我们考察两个基本的超循环H_A和H_B，并假定超循环H_A对H_B产生出某种催化生长因子，且反之亦然。这样一种生长因子也许是此超循环的一个组分或者是由它产生的某种物质。从我们先前的经验中，我们可以预期，相互的催化增长作用将导致协同行为。

表 14　关于催化耦合的超循环 H_A 和 H_B 的固定点分析

(1) 三级耦合项

$$\dot{c}_A = k_A c_A^2 c_B - \frac{c_A}{c_0}\phi;\quad \phi = (k_A c_A + k_B c_B) c_A c_B$$

$$\dot{c}_B = k_B c_A c_B^2 - \frac{c_B}{c_0}\phi;\quad c_0 = c_A + c_B \qquad (T14.1)$$

对此动力学系统的固定点分析得出：

$$\bar{x}_1:\ \bar{c}_A=c_0,\ \bar{c}_B=0;\ \omega=0 \tag{T14.2}$$

$$\bar{x}_2:\ \bar{c}_A=0,\ \bar{c}_B=c_0;\ \omega=0 \tag{T14.3}$$

此系统是竞争性的。分析更高级项（图 1-44）揭示了，对于 $k_A>k_B$，$\bar{x}_1$ 是稳定的。另一方面，条件 $k_A<k_B$ 导致了 $\bar{x}_2$ 的稳定性。

(2) 第四级偶合项

$$\dot{c}_A=k_A c^2_A c^2_B-\frac{c_A}{c_0}\phi;\ \phi=(k_A+k_B)c^2_A c^2_B$$

$$\dot{c}_B=k_B c^2_A c^2_B-\frac{c_B}{c_0}\phi;\ c_0=c_A+c_B \tag{T14.4}$$

这个动力学系统的特征是有 3 个固定点：

$$\bar{x}_1:\ \bar{c}_A=0,\ \bar{c}_B=0,\ \omega=0 \tag{T14.5}$$

$$\bar{x}_2:\ \bar{c}_A=0,\ \bar{c}_B=c_0,\ \omega=0 \tag{T14.6}$$

$$\bar{x}_3:\ \bar{c}_A=\frac{k_A}{k_A+k_B}c_0,\ \ \bar{c}_B=\frac{k_B}{k_A+k_B}c_0;$$

$$\omega=\frac{k^2_A k^2_B}{(k_A+k_B)^3}\cdot c_0^3 \tag{T14.7}$$

所以，$\bar{x}_3$ 代表了一个稳定的固定点，它表示在全部可能条件下两个耦合着的超循环 H_A 和 H_B 的协同行为。图 1-44 中显示的向量场表明，另外两个固定点 $\bar{x}_1$ 和 $\bar{x}_2$ 都是发点。

为简化直接分析，我们假定在两个超循环中都已建立起内平衡。催化项在分子浓度方面是三级的（分别为$k_A c_A^2 c_B$和$k_B c_A c_B^2$，参见表1-14）。因而在充分高的浓度条件下，我们可以忽略掉这个非催化系统的二级生长函数。固定点分析尚不足以研究所获得的动力学系统，因为它对所有简正模产生出零本征值。不过，容易对此向量场进行考察，因为此系统只有一个自由度。正如我们从图1-44所见，虽然存在着催化因子，两个超循环将仍然竞争。因此，所引入的这种催化耦合尚不足以导致协同行为。

在生物大分子水平上是难以将这类催化耦合的物理现实性加以形象化的；在总的速率方程中存在象$k_A\ c_A^2 c_B^2$或者$k_B c_A^2 c_B^2$的某个项，或是需要复杂的多步机制，或是需要两个以上的大分

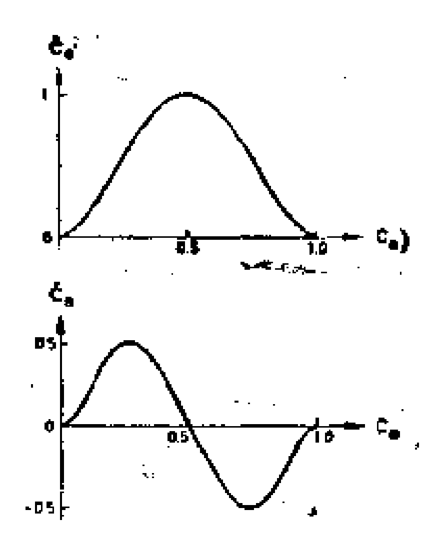


图 1-44 超循环之间的耦合 *a*)催化耦合项分别是 $k_A c_A^2 c_B$ 和 $k_B c_A c_B^2$。除了在两个固定点以外，在物理上允许的区域以内 ($0<C_B<1$)，切向量为正：假定 $k_B>k_A$，因而超循环 C_B 被选择。虽有耦合项，此系统仍然是竞争性的。*b*)催化耦合项分别是 $k_A c_A^2 c_B^2$ 和 $k_B c_A^2 c_B^2$。此系统含有两个在隅角上不稳定的固定点，以及一个在中心稳定的固定点($\bar{c}_A=\bar{c}_B=0.5$，因为 $k_A=k_B$)。此系统是协同的。

子发生碰撞，而这两者都是不太可能的。① 所以，人们受到启迪而得出结论：将高级催化项引入生长于均匀溶液中的系统，这样做似乎不会发展出更复杂的结构，这类结构由分级耦合的自复制单元组成，而是导致已经存在的功能单元间个体化发展。例如，在某个分隔间中将一个超循环中的所有成员进行空间隔离，就可以做到这一点。现在的细胞原型的形成，可以作为导致个体化超循环的一种可能的机制。在隔离起来以后，个体化的超循环，其行为可以象一个简单的复制单元。所以，超循环更可能是自组织的中间物而不是自组织的结局。

① 向以技术装置为基础的人工动力学系统中引入催化耦合项，例如电网，则不会遇到这些困难。

结论

第二部分的主要目的在于，对自复制系统中的各种功能连接进行抽象的比较研究。所使用的方法在微分拓扑学中是普通的。除了在一些特殊情况以外，往往得不到完全的分析解，因为所涉及的微分方程是内在地非线性的。自复制总是引起产生速率与相应物种的群体数相关。通过编码功能连接起来的不同物种中的协同作用，进一步强加上了浓度项，这导致了速率对群体变量的高级相关。

对选择和进化的行为进行比较分析，并不要求关于完全的解曲线的知识。为确定一个功能上协同的系综中的所有成员是否能够稳定共存，通常只要找到它们的最后结局就足够了。借助于李雅普诺夫方法，有些情况下还要更详细地考察完全向量场，固定点分析在此是相当有用的。综合分析的结果可以总括如下：

一个由几个自复制单元构成的系综，其功能整合要求在所有成员之间引入催化连接。加在亚单元个体复制循环上的这些连接，必须形成一个封闭环，通过所有群体变量的相互控制以使系综稳定化。在一定空间条件和有限的时间间隔中，可以存在于"小生境"以及催化链或者分支网络中的独立竞争者，缺乏作为超循环特征的自组织性质。为使系综中的所有成员相干地生长和进化，只有共存是不够的。特别是，超循环具有如下的显著性质：

1．它使得通过循环连接而联系起来的所有物种稳定地、受控地共存。

2．它允许它的所有成员相干地生长。

3. 此超循环与任何不属于此循环的单复制单元竞争，而不论那种实体是独立的还是另一个超循环的部分，甚至也不论它是否是被"寄生耦合"连接在此特定循环上的。

4. 超循环其大小可以放大或缩小，如果这种变更能提供任何选择优势的话。

5. 在更高级的网络中，超循环不容易连接起来。两个组织度为 p 的超循环，为了互相稳定而需要组织度为$2p$的耦合项。

6. 超循环的内连接和协同性质能够向优化功能进化。"表现型"优势即那些对突变体直接有利的变异立刻得到稳定化。另一方面，"基因型"优势，它们有利于后继的产物，因此只是间接地有利于其中出现了突变的复制单元；为了使竞争结果固定下来，这种优势需要空间隔离。

7. 一个超循环的选择，是一种"一旦则永存"的结果。在任一普通的达尔文系统中，提供了某种选择优势的突变体容易生长起来，并得到确立。它们的生长性质与群体的大小无关。对于超循环，选择优势总是群体数的函数，这是由于超循环的内在的非线性性质所致。因此，一个超循环，一旦建立起来以后，就不容易被任何新来者所取代，因为新物种总是作为一个（或n个）拷贝出现的。

所有这些性质使得超循环是一类独特的自组织化学反应网络。实质上这证明，更加正规地检验它们的性质是正当的，这也就是第二部分的目的。这类超循环的简单代表可在自然界遇到，这正如第一部分所表明的。这种类型的功能组织也是分布广泛的，并在神经网络或者在社会系统中起着某种作用。另一方面，我们并不希望把超循环当作某种偶像。它们在分子自组织中的作用是有限的。它们允许整合信息，犹如在翻译的起源中所需要的。

不过，只要具有高度复制精确度的酶机构建立起来，将整合系统个体化为活细胞形式，超循环就消失了。个体化的复制系统具有高得多的进一步多样化和分化的潜在能力。

存在着许多形式的超循环组织，它们可以是从直接的二级耦合直到 n 级的复合超循环，在复合超循环中，每一步反应都需要所有的成员协同行动。虽然我们并不知道有比二级超循环更简单的组织能够带来翻译器，但是我们清楚地意识到，哪怕是这种“可能的最简单的”系统也是很复杂的。所以，我们在第三部分的任务就是要表明，现实的超循环的确能够从原始条件下以充足丰度存在的更简单前体中出现。



现实的超循环

为“现实的超循环”而提出的模型，与某种原始复制和翻译器的分子自组织紧密相联。对于功能上相联系统的一切基因型表现型组分，超循环组织提供了选择稳定性和进化适应性。超循环组织起源于分子拟种中，通过突变和基因复制向更高的复杂性进化。它的早期结构看来反映在下述几方面中，即密码子对于氨基酸的分配，*tRNA* 的同源序列，复制和翻译的双重酶功能，以及原核细胞基因组的结构和功能组织。

3.1 怎样开始翻译？

“蛋白质合成的起源，是一个众所周知的难题。在此我们并不是指形成随机的多肽，而是指虽是粗糙的、但却是由核模指导的多肽合成的起源，现在的遗传密码其表达需要由活化酶、转移 *RNA*、核糖体、各种因子等所构成的复杂机构。”

这是从克里克、布伦纳、克卢格、皮森尼克最近的论文中摘录下来的，这段导言性的话贴切地概括了我们的主题。

让我们在此假定，某种粗糙的复制和翻译机构以合适的精

度起作用，并适应极其丰富的分子符号字母表，它们通过某种不必再进一步详加说明的(例如自组织或创生的)过程，而已存在于自然界中或实验室里。让我们进一步假设一种环境，这种环境提供了合成诸如核酸和蛋白质此类大分子所需要的一切活化高能原料，使复制和翻译成为自发过程，即是由正的亲和力所推动的。然而，这样出现的一个系综能象一个达尔文系综那样连续地进化吗？换言之，此种系综能够不断地把起初给出的信息积累起来，并不断加以改进直到取得最大的功能效率吗？

为了把这个问题落实到更为具体的情形，让我们考虑图1-45中所示意的模型。一组给定的 RNA 分子的正链包含了相应数目蛋白质分子的信息。翻译产物至少能够满足如下功能：(1)蛋白质的作用象 RNA 聚合酶，即类似于与各种 RNA 噬菌体相联系的专一复制酶。它的识别部位适应了出现在此种RNA的正链和负链中的某种特定序列或结构；换言之，它只能有效地复制那些证明了它们本身是特定系综的成员的 RNA 分子。

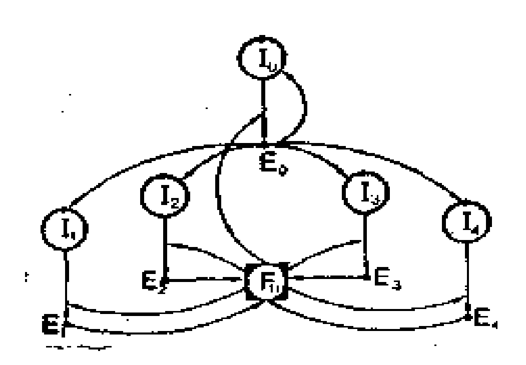


图1-45　最起码的原始翻译模型中包含了编码复制酶 E_0 的信使 I_0，它适应于专一地识别序列 I_0 到 I_4。I_1 到 I_4 的正链编码 4 种合成酶功能E_1到E_4，而负链可以代表 4 种氨基酸的连接物（tRNA）。这种系统虽然包括了翻译和自复制所需要的所有功能，但由于内部的竞争，它是不稳定的。如果 I_0 到 I_4 未被超循环连接稳定下来，相干地进化便是不可能的。

(2)其他翻译产物的功能象活化酶，它们将各种氨基酸独特地分配并连接到它们各自的 RNA 连接物上，每个连接物带有一个确定的反密码子。各种氨基酸个数以及不同的连接物个数，被调节到与出现在信使序列即 RNA正链的各种密码子相匹配，从而产生出具有确定密码子的"封闭"翻译系统。正如今天所知，它不必包含全部遗传密码，而可能仅限制于为数不多并具有足够功能的几种氨基酸(例如 4 种)，利用在密码子结构上的某些约束以保证某种明确的识别。连接物可以由RNA组分的负链来代表，若这过于局限则它们可以与更多的机构诸如核糖体一起被提供，这些机构处于恒定的环境因子形式之下，这类因子类似于细菌细胞内帮助噬菌体复制与翻译的那种宿主因子。初看起来，我们也许已对图 1-45 所示系统显示出高度的功能交织感到满意。其中所有 I_i得到复制酶 E_0的催化支持，E_0 反过来又将其存在归功于翻译酶 E_1到 E_4的联合功能 F_1，没有翻译酶则它不可能从 I_0被翻译过来。当然，酶 E_1到 E_4在产生出自身时也要利用这种翻译功能，但正因为作为 I_1到 I_4的翻译产物，它们最后也分别依赖于E_0或 I_0。然而，详细的分析表明，已有的耦合尚不足以保证不同的基因型组分 I_i的相互稳定化。由E_0实施的一般复制酶功能和一般翻译功能F_{tr}，在所有的微分方程中都由同样的项来代表。这些微分方程乘上一个普通的时间函数 $f(t)$，就还原为未耦合竞争者的微分方程。这个一开始作用起得相当好的系统，由于内部竞争，注定是要毁灭的。对此速率方程进行数值积分所获得的一组典型的解曲线如图1-46所示。图1-47是这种系统的另一个例子。在此所有的信使产生出它们自己的专一性复制酶 E_1到 E_4，它又提供了合成酶功能(F_{tr})。再者，由伴随的翻译功能带来的这种耦合，并不足以使此系综稳定。是否只要存在翻译功能与复制酶的信使系统以及翻译的产

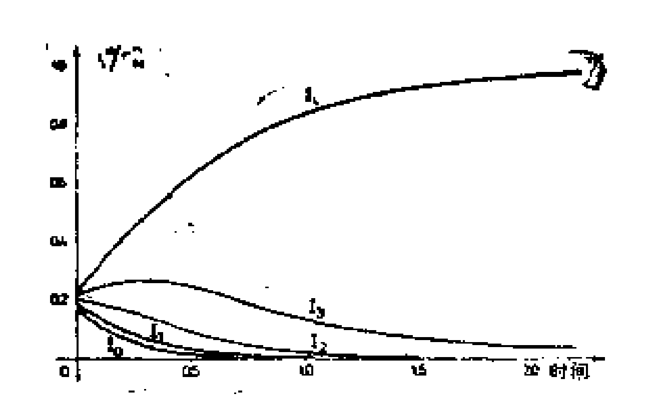


图1-46　类似图1-45所示模型的微分方程组的解曲线。在这个特例中，假定初始浓度和自催化复制速率常数从I_0到I_4线性地增加，而其他的参数，诸如翻译速率常数($I_i \xrightarrow{k_i} E_i$)、氨基酸分配（$E_1$、$E_2$、$E_3$、$E_4$对$F_{tr}$的贡献）或者酶-底物复合体的稳定性($I_i + E_0 \overset{k_1}{\rightleftharpoons} I_i \cdot E_0$)等等，对全部反应物都是一致的。相对群体数($y^0_i/C^0_N$)的时间过程反映了竞争行为。最有效的生长模板($I_4$)将取代所有其他模板而最终处于支配地位($y^0_4/c^0_N \to 1$)。不过，因为复制(由$E_0$代表)和翻译功能($E_1$、$E_2$和$E_3$对$F_{tr}$的贡献)，二者都消失了，$I_4$也将消亡。整个群体注定都要毁灭($c^t_N \to 0$)。

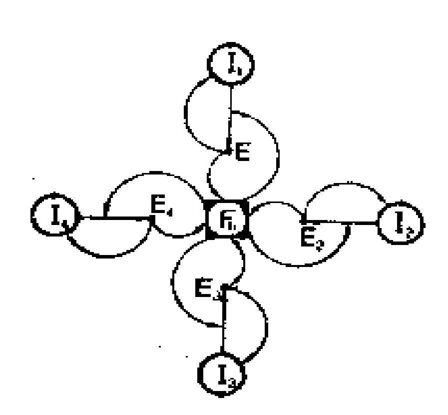


图1-47　在这个可供选择的关于原始复制和翻译的模型中，假定酶E_1到E_4都具备双重功能，即既作为它们自己信使专一性复制酶，又作为4种氨基酸分配的合成酶。此系统的命运与图1-45所示意的系统的命运是相同的，原因在于其中的信使是高度竞争性的。

物系统，就足以使它连续存在并进化？对于这个问题，我们的答案是，如果在各种复制组分 I_i之间不引入某种特殊的耦合，尽管它们包含了复制和翻译所要求的全部性质，这样的系统仍然是不稳定的。即使所有的组分是选择上等价的(或接近等价的)，并因此可以共存一段时间(取决于它们的群体大小)，它们也不能以互相控制的方式而进化，所以决不可能优化它们的功能相互作用。它们的最终命运总是毁灭，原因在于，如果它未被特殊的耦合所加强，则碰巧的选择上的等价不可能在较长的整个进化期间内相干地保持下去。

当然，已知第二部分的结果就不会对这种答案感到吃惊了。更深入地考察由复制和翻译酶的功能所提供的此类特殊的连接，并不揭示出任何超循环的本性。所以，这些连接不可能建立起群体数的相互控制，这种群体数是一个有组织的系统中各种组分相互联系而进化所要求的。在所研究的两个系统中存在的耦合，可以归结为两种普通的功能，这些功能就象环境因子一样以完全相同的方式影响所有的组分，因此不可能提供任何相互控制。

对于我们在本文所要表明的东西，上述例子是典型的，亦即：

1. 以低精确度的复制和翻译以及低起始丰度的有效复制为特征的进化初期阶段，与其他任何(结构的)组织相比，超循环组织具有较大的选择优势(参见 3.5 节)，以及

2. 超循环模型的确能够建立起来，从而给现在的原核细胞中发现的复制和翻译器提供了现实的前体(参见3.6节)。

给定了可能存在的象 tRNA一般大小的可复制RNA分子和合成原始形式蛋白质的前提，利用有限数目的(很普通的)氨基酸，我们如何设想翻译的起源呢？

3.2 原始编码的逻辑

3.2.1 RRY码

最近提出的关于模板指导蛋白质合成的推测性模型是十分诱人的，这个模型的基础是与无逗点和相干读出问题相联系的一些逻辑推理。一种原始密码必须有一定的框架结构，否则不可能一致地读出一条信息。偶然的阶段性差错将引起关于信息的部分翻译框架错位，从而破坏了信息的意义。此模型的提出者于是提出了一种特殊的碱基序列，所有的密码子都必须附着在此序列上。换句话说，只有那些类似于此特殊模式的核苷酸序列，才能承担信使功能。通过暴露的tRNA 反密码子环以及内部自复制所赋予的指令，可产生模式的均一性。他们在那些保证非交叠读出的可能模式中，选择了碱基序列嘌呤—嘌呤—嘧啶，一般记作 RRY，认为这对于说明一条信息的所有密码子都是共同的。这种特殊的选择，由于在现在的 tRNA反密码子环中发现了一种序列规则性，从而得到了支持。这种序列记作 3′NR$\alpha\beta\gamma$UY，$\alpha\beta\gamma$ 是反密码子， N是 4 种核苷酸中的任何一种， R和 Y 分别是嘌呤和嘧啶。由信使和生长的核苷酸链所形成的复合体的稳定性，是无核糖体的翻译的另一个前提。在转移到随后的氨基酰 tRNA完成之前，即在翻译完所有信息之前，肽酰 tRNA不可脱落。否则只会获得没有功能效率的蛋白质片段。从已知的碱基对的稳定性来看，简单的密码子-反密码子相互作用显然不足以保证信使 tRNA 复合体所需要的稳定性。因此，此模型主要是以 3 个辅助假定为基础。

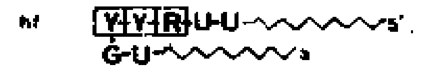
1. 连接物（*RNA*前体）的反密码子环结构是这样的：在给定了特定的、共同的密码子模式后，一个RNA总能与信使形成5个碱基对。于是，原始 *tRNA* 分配到一般的反密码子环序列

3′～～～～U-G-Y-Y-R-U-U～～～～5′

式中 YYR是反密码子。

2. 可以假定，每一个原始tRNA的反密码子环都有两种不同的构象①，细节见图1-48。沃西在一篇早期论文中已描述了两种构型，他把它们称作 FH和 hf（FH指的是富勒和霍奇森，他们最先提出，反密码子环中未配对的7碱基序列的3′端上的5个碱基堆积在每一个的上面，而 hf是沃西提出的，它表示一种互补构型，这种构型让环的5′一端的5个碱基保持在某个堆积位置中）。沃西假定，两种构型之间的转变在核糖体蛋白合成中起着重要的作用，但也涉及到它过去的即更原始的机制中的可能意义。

3. 实际上出现其中哪一种构型，要取决于tRNA是附着在一个氨基酸上还是附着在一条肽链上。通过把一个氨基酸转移到此条肽链上，则 tRNA从 hf构型翻转为FH构型（参见图1-



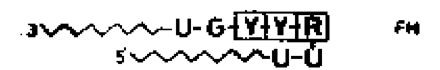


图1-48 *tRNA* 反密码子环的两种可能构型：按富勒和霍奇森为 FH，按沃西为 hf。（所构造的）反密码子模式是克里克等人的模型。

① 原文如此。——译者注

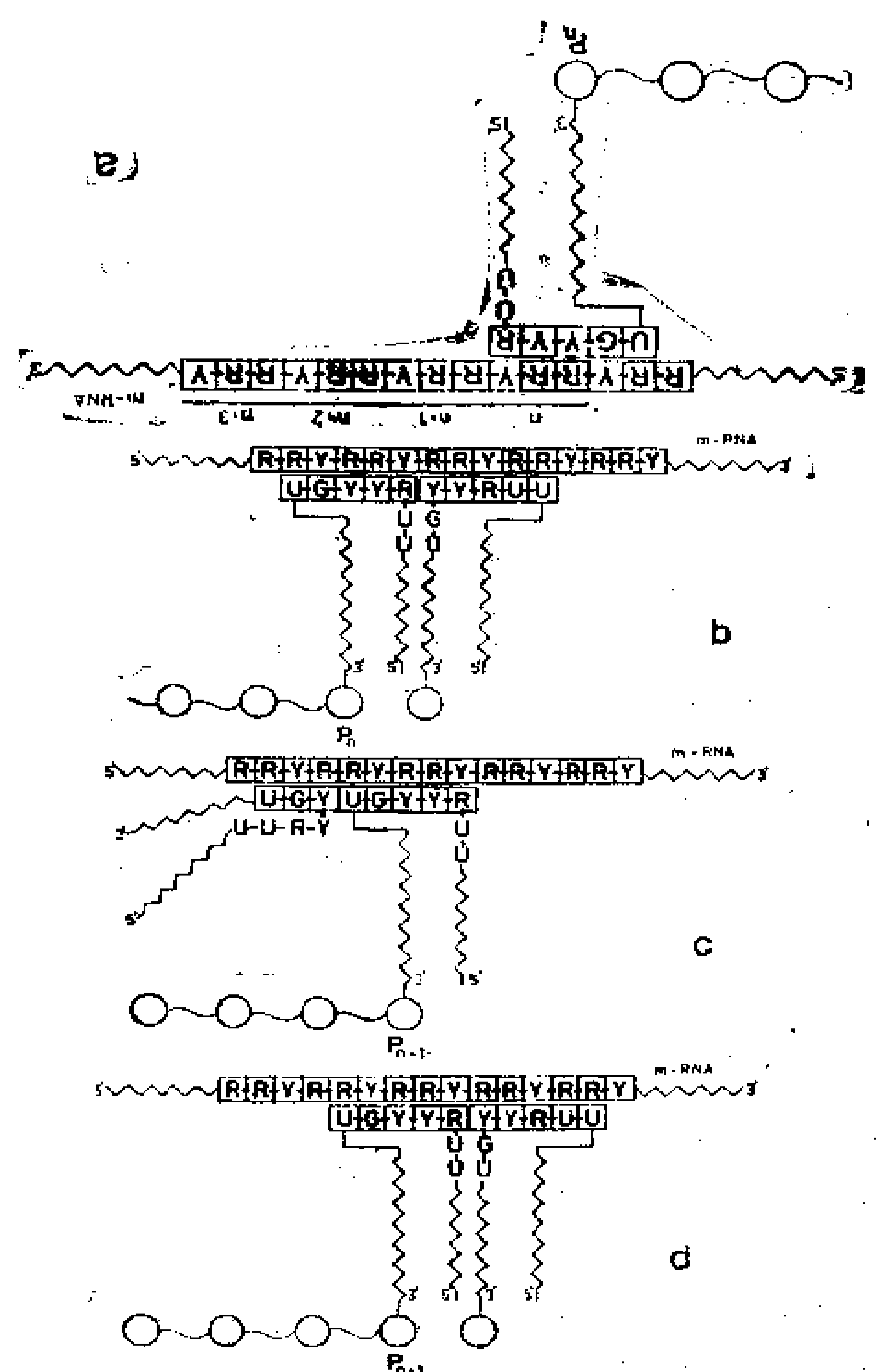


图1-49　原始翻译机构需要信使和肽酰 tRNA 之间的"粘性的"相互作用，从而允许生长的肽链与信使保持接触，直到完成翻译。根据克里克等人的看法，涉及 tRNA 构型变化(FH⇌hf)的翻转机制会影响输运。新生肽链通

过 5 个碱基对总是与信使相联系。由于邻近的氨基酰 tRNA 而获得一些附加的稳定性。碱基配对的部分交叠，保证了一致地阅读碱基三联体编码的信息。

49)。

另外，还有第 4 个假设，虽然作为此模型的前提它并不是绝对需要的，但是，它引起信使上的两个相邻 tRNA 之间的某种相互作用，这种相互作用保证了所需构型将占有最大优势，并对多肽-信使复合体的稳定性有进一步的贡献。

图 1-49 更详细地表明了，以给定的论据为基础是怎样促进了多肽合成的。运用转移反应的自由能作“燃料”，生长的多肽链沿着信使被输送。这个反应的进行可以借助于一般的非专一催化剂，但是此机制倒是不要求诸如由现在的核糖体所提供的那种高级机构。尽管 5 个碱基对稳定了密码子和反密码子之间的相互作用，主要的是实际上密码利用了碱基三联体。以前已有人强调过，不利用 3 个碱基的原始密码是无法解释现在的密码的。此模型则很好地解释了，在没有高级翻译器时，信息是怎么能够以如下方式读出的：

自发地，

相继地，

完整地，即不是片段地，以及

可复制地，即严格地保持给定的密码的框架。

密码内在地与现在的 tRNA 反密码子环的一些结构特征相联系，这意味着这些分子是最初的有功能组织实体的子代。由这个模型所规定的 4 种氨基酸是：

GG_C^U；	GA_C^U；	AG_C^U；	AA_C^U
甘氨酸	天冬氨酸	丝氨酸	天冬酰胺

其中的一些氨基酸在原始汤中是十分丰富的。

另一方面，此模型也带来了一些困难。一个一致的 RRY 序列，其中嘌呤大大多于嘧啶，因此难以有利于稳定的内部折叠。

结果是这样的序列：

对于水解是相当不稳定的(如果以单链存在)，

极其倾向于形成双螺旋(原始的翻译机制难以对其进行复制)，

缺乏内部对称，并且

产生出另外一种一般的密码模式(即 5′RYY)的负链。

3.2.2 RNY码

在对上述各点进行更详细的讨论之前，让我们在此提出另外一个模型，它没有上述特殊困难。在*RNY*中，N代表 A、U、G、C 其中的任何一种。这个模式的提出，也应归功于克里克等人。不过，克里克等人不喜欢它，原因在于一种不利的背景：如果N代表某种嘧啶，那么，具有一般序列 3′ⓊGYNRUⓊ的反密码子环，在某些情况下只能运用在它中央的 5 个核苷酸去与信使形成稳定碱基对。不过，观察事实否定了这个论据；观察表明，一个 RNY 码能够分配到 8 个氨基酸，从而人们可以排除掉那些不能满足信使-肽酰 tRNA 复合体稳定性要求的组合(参见后面)。

一般化的密码模式RNY的优点何在？首先，RNY 码象其类似物 RRY 一样，是无逗点的。而且，它对于正链和负链是对称的。如果从 5′记到 3′位，则正链和负链的一般框架结构都是 $R^{N}_{N'}Y$。这里，N和 N′ 是互补的，并位于两链序列中的

镜象位置上。类似的对称性也能够在一链的内部发展起来，从而允许形成对称的二级折叠结构。(几乎是)对称折叠的一个典型例子是现存的tRNA。单链噬菌体基因组及其衍生物(诸如QβRNA的中变体)也有这样的对称元素作为显著特征。对称结构的选择优势在此是明显的。如果这些分子要由专一地识别某种结构特征的聚合酶复制出来，那么只有某种对称结构才会允许正链和负链成为等效模板。这种效率的等价性是选择所要求的。因此，*tRNA* 的对称性很可能是这些分子仍必须自发复制的时代的一种遗迹。

内折叠还有利于这些分子抵制水解，并提供指令正确读出信息的一条方便途径。在开放结构中，其中许多核苷酸保持未配对(例如，对于RRY序列)，复制和翻译可能始于此序列的任一未配对位置，从而导致了片段产物。在已完全配对的结构中，未匹配的两个粘性端注定要成为复制和翻译的起点。以这种方

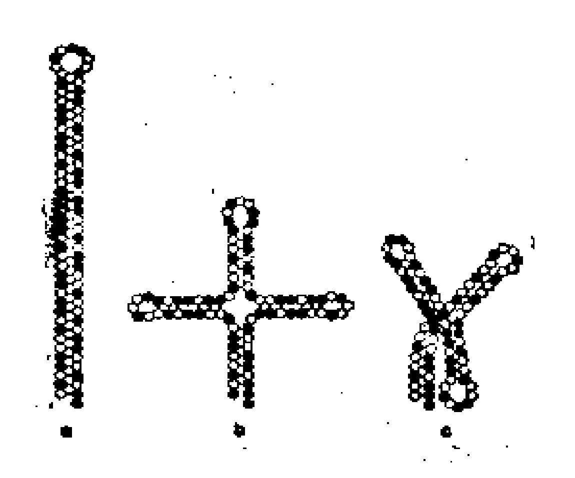


图1-50　RNA的对称二级结构要求相应的内部互补性(参见第一部分的图1-14)。正链和负链则都将显示出相类似的折叠。RNY密码子模式能够产生出这样的结构。已设计了一种博奕，它的规则考虑到了所观察到的低聚核苷酸和tRNA的物理相互作用。此博奕表明了最可能出现的结构是哪一种结构。发夹型结构需要两个互补的半个分子，例如，连接起来的一个5′RRY序列与其含有一个5′RYY模式的负链。

式可读出完整信息，它只要求模板结构暂时的部分折叠，它可能被与生长着的链相互作用强制着执行的。对称性虽然不是绝对需要，但是在这种情况中它是有利的(参见图1-50)。

从逻辑观点来看， RNY码看起来比 RRY 码更有吸引力，这在于以下 3 个论据。

1．RNA分子的选择性增强，必须对正链和负链都有效。对称的 RNY模式比起RRY负链更容易能够满足这种要求，因为两者的负链(*RNY*与 *RYY*) 不同，因此两者不可能成为酶专一识别的等价目标。

2．从高度复杂性的观点看，找到对于复制和翻译提供有用性质的很少几个序列的机会是非常小的，如果这些序列由于其对称结构，满足(1)中所列的要求，那么正链和负链都可以是体现这类功能的候选者。

3．翻译器及其各种 tRNA和信使的进化，要求所有的复制分子有相互稳定作用。正如稍后将表明的，这是由于其对称性能提供两种互补的功能，从而超循环从拟种中出现要容易得多。

一个系统，只有它采取了选择性自组织的“达尔文逻辑”，它才能够积累信息并最终进化到更高的复杂性。不过，这种逻辑必须在物质性质中找到它的基础和表达方式。开始时这些组分所能认识的，不过是自然丰度和相互作用强度。上述性质，就是我们为了理解出现翻译所必须要加以讨论的。

3.3 原始编码的物理学

3.3.1 开始的条件

作为多分子过程的自组织，要求单体以及聚合体以充分的浓度存在。所以，开始自组织之前，必须有一个前生物合成的扩张阶段，在这个阶段中积累了为产生"高度富集的汤"所必需的所有材料。我们不打算停留在这些原始化学过程，也不打算卷入关于历史边界条件的争论之中。如果要估计历史过程的绝对速率，那么诸如"汤"是否起源于海洋中、池塘中或者甚至小泥潭中，或者是不是涉及到分界面、粗纹理的或多孔的表面等等问题，都可以是重要的。

在此我们只从如下假定出发：当自组织开始时，所有各种高能原料都是普遍存在的，尤其是存在着：

各种丰度的氨基酸，

包括 4 种碱基*A*、*U*、*G*、*C*的核苷酸，

前面这两类聚合物，即类蛋白质以及类 *tRNA* 物质具有或多或少的随机序列。

在此所谓"较少的随机"指的是存在着紧邻的、更复杂的折叠相互作用，这导致某些结构的优先权。而"较多的随机"指的是，它们原先与任何预定的功能目标无关，如果起初存在着这种预定目标，那也只是巧合。

另一方面，我们肯定不能假定：已存在着诸如专一的复制酶，适应了的合成酶，或者任何一种核糖体功能那样的已适应了的蛋白质机构。

这并不排除非指令性低适度的蛋白质催化剂对开创复制和翻译起的促进作用。不过，那些蛋白质不能复制和有选择地改进，在“恒定的环境因子”条件下必须把它们与其他催化表面划入同一类。

3.3.2 核苷酸丰度

由于只有类核酸结构才提供了内在的自复制性质，所以首先更详细地分析一下它们的丰度以及它们的相互作用是重要的。

单体核苷酸尤其是它们的高能寡磷酸酯形式，远比氨基酸难以获得(运用可能的前生物机制)。因此，很少见到关于相对丰度的定量陈述。米勒和奥格尔强调腺嘌呤核苷酸在遗传过程以及在能量转移中的主要作用，并把它与这种物质比较容易形成相联系。奥罗及其同事发现，在浓的氰化铵水溶液中可以获产率为0.5%的腺嘌呤，而米勒和奥格尔表明，在由阳光催化的反应中，甚至仅仅氰化氢也产生出重要的中间物：

$$4HCN \longrightarrow \text{四聚物} \xrightarrow{h\nu} \begin{array}{c} N\equiv C \quad\quad N \\ \| \\ H_2N \quad\quad NH \end{array} \longrightarrow \text{腺嘌呤}$$

这样的中间物还能够与氰酸盐、尿素或氰进行反应，得出鸟嘌呤。对嘧啶合成机制的了解要少一些。运用甲烷和氮的混合物的放电产物丙炔腈与氰酸盐化合，可以演示合成胞嘧啶的一种途径。另一方面，尿嘧啶看起来象是胞嘧啶的水解产物，最初出现也可能归于这种来源。

关于嘌呤和嘧啶的原始自然丰度，我们讲不出多少东西。

模板指令聚合的速率，与所涉及到的单体浓度成正比。对于互补指令，至少需要两种核苷酸，它们在信息序列中必须是等价表示。所以，低丰度核苷酸的内含物是限速步骤，至少对于链的延伸是这样。于是，除了低聚A引物的聚集是限速步骤的情况以外，原始单体分布中可能出现A大大超过U，这几乎无助于偏爱共聚物AU而不是GC。复制生长的能力受到丰度较小的互补核苷酸对的模板功能的限制。在原始条件下，如果G和C的丰度处在A和U之间，那么也许会以相当的速率很好形成富GC和AU的共聚物。

所以，我们不能坚持此种较早的、推测性的观点：最初的密码子是从一张二元字母表独特地被征募过来的，这张字母表单独由AU组成。

3.3.3 互补结构的稳定性

为更好地阐明这个最初的密码子问题，碱基对的稳定性可以提供一些线索。碱基配对的稳定性和速率，我们已用各种核苷酸组合对其进行了研究。这些结果在早期的评论中已详细讨论过。它们定量地揭示了普遍接受的观点，即认为在协同堆积中 GC 对提供了比 AU 对大得多的稳定性。

一个连续均一的 n 个核苷酸对的低聚序列的稳定常数，可由如下关系表示：

$$K_n=\beta s^n \tag{91}$$

这涉及到线性的伊辛模型。因子 β 是协同性参数，它对于AU对和GC对两者等于10^{-3}的数量级①，而s 是在协同堆积中的一

① 这种关系，对于给定序列中的内部碱基配对和两个互补序列的二元缔合，在形式上都是有效的，其中 β 的量纲为M^{-1}。

个配对的稳定常数。对于 AU对的均一序列，这个参数 s 约比 GC 对的均一序列小一个数量级，或者大致定量表示为：

$$s_{AU} \approx 10 \qquad \text{当} \ s_{GC} \approx 100\text{时}$$

如果互补链中的一条链能够在 tRNA的反密码子环中采取特殊的堆积构型，则发现绝对稳定性比关系式(91)所预言的更高。可以推测在这种情况中协同参数 β发生了变化。然而，乌伦贝克、巴特和多蒂发现，与tRNA的反密码子区域互补的，只有一个 AU对不同的 3 聚核苷酸和 4 聚核苷酸，其稳定常数表现出有一个数量级的差异，这与上述数字是相当一致的。可以合理地认为，所发现的稳定常数的最大绝对值，是其反密码子中互补的两种 tRNA相互作用的稳定常数。

在一定的短序列中所获的这些数据，至少可用来对各种复制和翻译进行比较，并就其相对意义作出结论。显然，在任何现实浓度条件下，单独分离出来的 AU或 GC 对是不稳定的。因此，开始复制需要得到引物形成形式的特殊帮助，特别是这一步首先需要酶的支持。现在的噬菌体 RNA 复制酶，也是专一地适应了噬菌体基因组的一级序列模式。为了链的延伸，新来的核苷酸被协同地约束在此生长链的碱基对堆积的顶部。在此，这些数据提示：GC对大约比 AU对稳定10倍，导致 G和 C的精确性 q 相对地比 A和 U 的高。如果复制速率由聚核苷酸骨架中共价链联接的形式(而不是由形成碱基配对)所制约，那么可把精确度与单体浓度 m_R和 m_Y以及配对稳定常数K_{RY}、K_{RR}和 K_{YY} 联系起来。于是，任何给定的核苷酸的复制精确度可从这两个互补过程的精确度的几何平均值求得，

R⟶Y； Y⟶R：

$$q_{RY} = \frac{m_Y K_{RY}}{\sum_N m_N K_{RN}} \quad \text{及} \quad q_{YR} = \frac{m_R K_{YR}}{\sum_N m_N K_{YN}} \tag{92}$$

式中 $K_{RY} \approx K_{YR}$ 对所有的 N=A，U，G，C求和。那些 q 值分别对于 A和 U或者 G和 C 是一致的，因为误差可以出现在正链中，也可以出现在负链中。如果单体浓度的大小是相等的，则稳定常数决定了可以获得什么样的精确度。结果发现，G 和 C的复制比 A和 U的复制要精确得多。不过，在混合系统中GC和AU 的复制误差率的比，并不完全类似于相应稳定常数的(反)比，主要原因在于存在着 GU 摆动相互作用，这是复制误差的主要来源，即使在现在的RNA噬菌体复制中亦如此。

我们以无酶的核苷酸相互作用的各组数据为基础推测了 q 值，结果总结在表 1-15 中。前面的 3 组数据系指A、U、G 和 C 的单体浓度相等。这种假定可能是不现实的，因此在后面的 3 个例子中作了修订。人们也许会拒绝应用这些从低聚核苷酸的研究中所获得的稳定性数据。不过，一个核苷酸参与复制过程包含着协同的碱基对相互作用，因此应该相似于在低聚核苷酸中发现的有关数量级。对于计算 q值所需要的，是相对的稳定性，而不是绝对的稳定性。

从表 1-15 中提出的各种估计值所得的结论是：G和C 的复制精确性显著地高于A和U。依赖于被选择序列的优势[σ，参见第一部分(28)式]，在早期复制过程中富 GC 序列的可复制信息量，被限制在大约20至100个核苷酸即相当于类 tRNA分子，而富 AU序列则很难超过每个复制单元有 10 至 20 个核苷酸残基。在这里应强调的是，对于任一种组分，也许都已出现了较长的序列。不过，它们是不可复制的，从而是不可能依据任何功能的需要而进化的。

对噬菌体复制的实验数据进行分析，使我们在第一部分得出结论：哪怕是已很适应了的RNA复制酶，也不会允许每链上可复制地积累起超过1000到10000个核苷酸。这相当于RNA噬

表1-15　G和C与A和U的复制的误差率和精确度的估计

单体浓度	碱基对的稳定常数	精确度q		误差率$1-q$	
		GC	AU	GC	AU
$m_A=m_G=m_C=m_U$	$K_{RR}=K_{yy}=1$ $K_{AC}=1$，$K_{GU}=10$ $K_{AU}=10$，$K_{GC}=100$	0.93	0.59	0.07	0.41
$m_A=m_G=m_C=m_U$	$K_{RR}\approx K_{yy}\ll 1$ $K_{AC}=1$，$K_{GU}=10$ $K_{AU}=10$，$K_{GC}=100$	0.95	0.67	0.05	0.33
$m_A=m_G=m_C=m_U$	$K_{RR}=K_{yy}\ll 1$ $K_{AC}=1$，$K_{GU}=5$ $K_{AU}=10$，$K_{GC}=100$	0.97	0.78	0.03	0.22
$m_A=10m_G$ $m_G=m_C$ $m_C=10m_U$	$K_{RR}\approx K_{yy}\ll 1$ $K_{AC}=1$，$K_{GU}=5$ $K_{AU}=10$，$K_{GC}=100$	0.93	0.81	0.07	0.19
$m_A=10m_G$ $m_G=m_C$ $m_C=10m_U$	$K_{RR}\approx K_{yy}\ll 1$ $K_{AC}=1$，$K_{GU}=5$ $K_{AU}=10$，$K_{GC}=100$	0.95	0.69	0.05	0.31
$m_A=10m_G$ $m_G=m_C$ $m_C=10m_U$	$K_{RR}=K_{yy}=1$ $K_{AC}=2$，$K_{GU}=10$ $K_{AU}=10$，$K_{GC}=100$	0.86	0.25	0.14	0.75

菌体的实际基因量。

我们现在可以对原始复制机制作出完全的陈述：只有对于富GC结构，tRNA那样大小相当的序列才是可复制的。因此：

富GC序列有资格候选为早期的tRNA连接物和可复制的信使，至少对于不借助于恰当适应的酶的复制是这样。

对于开创翻译，可以得到类似的结论。正如克里克等人强调的，肽酰tRNA–信使复合物的稳定性，对于任何原始的翻译

模型都是关键性的。运用上面所引的数据，一个包含了5个GC对的复合体的稳定性常数相当于：

$K_{5GC} \approx 10^{7} M^{-1}$

而对于5个AU对则要低5个数量级：

$K_{5AU} \approx 10^{2} M^{-1}$

再者这些值必须被看作是相对的；如果涉及到堆积环区域或$tRNA$（正如我们现在所认识到的），则它们实际上要稍大一些，不过，这无损于以相对大小为基础的论据。

我们还能够以寿命数据为基础去评价此模型。业已一致地发现，对低聚核苷酸互补链所测定的重组速率常数，其数量级在：

$k_{R} \approx 10^{6} M^{-1} S^{-1}$

给定了上述稳定常数，复合体的寿命将分别为

$\tau_{5GC} \approx 10s$ 及 $\tau_{5AU} \approx 10^{-4}s$

如果原来的稳定性更高一些，以及如果两个附着在信使链上的相邻的$tRNA$能够互相稳定，那么这些数据又可以变得大一些。于是，寿命正好满足富GC序列开创原始翻译。如果AU对过量，这些寿命必定太短了。我们现在明白，RNY码比RRY码稍为不利一些，这是由稳定性引起的，这种不利条件主要可以利用G和C，至少在一部分R及Y位点上面加以平衡。一个4元GC结构，肯定要比包含了多于两个AU对的任何5元结构更为稳定。

结论是：

tRNA前体和信使的富GC结构都大大有利于开创翻译。

3.4 GC配对密码

3.4.1 最初的两种密码子

如果我们把从稳定性数据所得出的结论，与由克里克等人提出的论据结合起来，我们就可以预言哪一些密码子规定可能是最早的。

只是那些其中G和C残基占支配地位的序列，才能够忠实地复制它们自己的足够长的序列。最初的密码子则只是这两种残基的独特组合物。无逗点读出的要求，把对称的组合GGG/CCC和GCG/CGC排除在外。写出这样的序列，容易验证这一点。具有正确的反密码子组合的连接物，能够结合在各种交叠位置上。如果引入了从对称前体衍生出来的更多密码子组合，则将得到更有害的结果。因此我们有两个组合的互补配对，即GGC/GCC和CCG/CGG(全部模式都从5′位读到3′位)。从对称性观点看，两者显得是完全等价的。不过，以第3个位置上的摆动为基础，则稍有些不对称。让我们比较由CNG或GNC密码子独特地构成的信使序列。在第1种情况中，摆动碱基G总是位于第3密码子位置，而在第2种情况中，它就被限制在第1个位置上，在正链和负链中(如果一致地从5′位读到3′位)都是一样。对于复制，不同的密码子位置是不可区分的。因此，无论何处出现了某个摆动碱基，在互补链中就会引入两种解释，当开始翻译时，在一种情况下影响第1个密码子位置，在另一种情况下影响第3个密码子位置：

5′……CNG　CNG　CNG……3′

↓

5′……${}^{C}_{U}$N′G${}^{C}_{U}$N′G${}^{C}_{U}$N′G……3′

以及

5′……*GNC GNC GNC*……3′

↓

5′……GN^{C}_{U} GN^{C}_{U} GN^{C}_{U}……3′

只有在第2种情况中，复制序列才被正确地翻译，即倘若与连接物的摆动相互作用优先出现在第三个位置上而不是在第一个密码位置上。

换言之，带有反密码子3′*CNG*的连接物，既可读作5′*GN′C*又可读作5′*GN′U*，而带有反密码子3′*GNC*的连接物，只可读作5′*CNG*而不能读作5′*UN′G*。如果为了保持连接物连接在信使上需要5个碱基对，则这种论据也许是无力的，这一来便无法清楚地区别摆动的位置。不过，在第1个和第3个密码子位置之间的这种不对称关系的确是存在的，在现在的遗传密码中是明显的。① 相对小的选择优势往往就足以偏倚进化过程。克里克等人显然也基于这样的论据而倾向于*RRY*(或*RNY*)模型。

我们现在能够对于最初的两种密码子作出唯一的规定，即

5′GGC　和5′GCC

如果以反平行的方式安排，它们就是互补的。这种选择可由4条论据来说明，即：

连接物-信使相互作用的稳定性，以及

复制忠实性，这两者都有利于GC　组合的开始，

需要某种非对称GC　模式的翻译中的无逗点读出，还有

翻译的一致性把摆动的双关性限制在第3密码子位置上。

我们将强调指出，这些论据只基于核酸的性质。令人满意

---

① 在定常分布中，*G*比*C*更为持久，这一事实有利于我们的论证。

地注意到，在现今遗传密码中的两种密码子GGC和GCC与两种最简单的氨基酸，即甘氨酸和丙氨酸有关，它们在模拟原始条件的实验中，的确以非常大的丰度出现。

人们可能提出反对意见，只由这两种残基构成的翻译产物将很难代表任何高级的催化因素。我们将在16节再回到这个问题上来。此刻只要注意到，在这个阶段翻译还不是保存潜在信使所需要的性质。单独选择最初的富GC链，其基础只在于结构的稳定性及它们的忠实复制能力。许多不同的GC 序列都同样好地适用于这个目的，因此也许已经被一起选择为(或多或少简并的)一种拟种的成员。对称的结构在此是大大有利的，因为它们能够同时满足正链和负链稳定性的标准。

在稳定的成员之中，可能已出现了无逗号码序列并且开始了翻译，这也许是由反密码子环的模板功能和后来的模式复制所引入的。如果它们的翻译产物对它们的信使的稳定性或者复制速率有利，那么它们将通过达尔文机制进一步进化，从而连续地改变此拟种的分布。在我们返回到这种借助翻译产物的稳定性之前，让我们再扩大一点关于结构稳定性与复制效率的问题，因为似乎所需要的两种性质是以互相冲突的前提为基础的。

3.4.2 “非周期的线性GC晶格”

具有内部折叠结构的类*tRNA*分子由氢键所加强，这可被看作一种微晶。如果它包含了延伸较长的互补GC模式，则所形成的内部结构可以是相当惰性的。从tRNA环或相应的低聚核苷酸的熔化曲线可知，只有4个GC对的完整序列就已相当稳定了。科茨研究了一种从酵母取得的低聚核苷酸，它相应于$tRNA^{Ser}_{1,2}$的外环，其中含有4个GC对(可通过tRNA分子的部分消化而

获得)。他发现，其熔解温度为84 ± 1°C，ΔH为44 ± 4千卡/摩尔。这相当于在25°C时的稳定常数为2×10^5，与前面讲到的数值相当一致。对拟种选择的评价不仅仅是这些链的结构稳定性，还有结构稳定性与复制效率的某种有利的组合。有效的模板功能要求一个环能迅速地部分解开，在这个过程中不允许GC对伸展得过分长。不过：

自然的序列无论如何都不是完美的。

给定高丰度的A单体和碱基配对的有限精确度，则GC微晶中总是掺入了很多A残基，其行为就象线性GC晶格中的缺陷。据推测，也许任何一种序列都从多到少地含有*A*、*U*、*G*或*C*。将被选择的并从而是复制增殖的序列不是完美序列，而是富GC序列。例如，在这种富GC序列中每个第5号位置都由一个A或U残基取代，于是，依据内部互补性的碱基配对区将平均含有不多于4个GC对(参见现在的tRNA)。那些结构，特别是在复制过程是借助于蛋白质时，容易发生区域熔解，所以这是代表了最原始形式的复制酶。

注意：在非周期性GC晶格中的A—U缺陷是选择上有利的。

正如托马斯·曼在《魔山》中所说："生命从绝对完美中倒退。"

3.4.3 从GNC到RNY

在被选择拟种的富*GC*链中出现一定丰度的*A*(以及互补的*U*)缺陷后，密码进化的下一步似乎是预编顺序的。突变可以出现在3个密码子位置中的任何一个位置上，但是它们的结果差异很大。取代密码子中央的碱基，将促进负链上相应的反密码

子中央碱基互补性的取代作用，从而立刻引入了两种新的密码子GAC和GUC。另一方面，在第1位或第3位上的变化，将分别由负链上的第3位（或第1位）上的变化来互补，依据摆动论据，这终将导致只有一种进一步的分配。而且，用于无逗点阅读的GC框架将受到干扰。

稳定性的要求，本来并不允许在信使-tRNA复合物的5个碱基对区域中取代一个以上的AU对。因此下一个最可能出现的密码子是5′GAC和5′GUC。作为预先存在的配对5′GGC/5′GCC的突变体，它们可以作为被选择的富GC拟种的成员而以高丰度出现。然而，如果这些突变体被分配给某个翻译功能，它们必须变得真正与占统治地位的5′GGC/5′GCC种相当。在这个阶段，4个密码子连接物（以及编码耦合因子的信使）的超循环稳定性成为绝对必要。没有这样的连接，原始翻译系统的不同成员虽可以共存一段时间，但是它们决不可能以任何相干的形式进化或优化它们的协同作用。

4种密码子允许4种不同的氨基酸分配，于是可以提供一种功能丰富的调色板。因此，所形成的蛋白质可以变成有效的耦合因子。作为同一种拟种成员的信使和tRNA，也许已出现在同一种RNA物种的互补链中，从而共有两种功能。

另一方面，对于它们的继续进化来说，这可能是一种太严格的约束条件。于是我们必须假定，它们都是从某种共同前体衍生出来的，但由于它们结构和功能的要求大不一样，所以后来分化成了不同的序列。

依照现在的遗传密码表，GAC和GUC编码天冬氨酸和缬氨酸。在更详细地讨论氨基酸之前，让我们粗略浏览一下进一步向更普遍密码进化的步骤。

对密码子-反密码子相互作用的高稳定性要求，随着翻译

产物越来越适应而不断降低。最后采取了摆动相互作用，GC框架码能够向更普遍的RY框架码进化。所有这些合在一起，导致出现了4种以上的氨基酸。第一种取代仍然是在稳定性约束条件下出现的，这迫使AU的含量尽可能低。因此，它引入了两种密码子，即5′AGC(=丝氨酸)和5′ACC(=苏氨酸)。它们的互补序列对第3个密码子位置产生影响，产生出5′GCU和5′GGU，从而复制出对丙氨酸(GCC)和甘氨酸(GGC)的编码。简并性被用来解释这些后来的密码子复制中的摆动相互作用，它也许是出现AGC和ACC密码子及其分配的最初原因。

如果随着酶机构的进化已允许在密码子区有1个以上AU对，我们又得到了两种新分配，即AA^{U}_{C}(天冬酰胺)和AU^{U}_{C}(异亮氨酸)。这就完成了RNY码的所有可能的分配。遗传密码的进一步进化，需要减弱非交叠框架的约束条件。所以，核糖前体的适应性现在就是绝对必要的了。

3.4.4 氨基酸的原始字母表

对各种氨基酸的原始丰度能够作出相当可信的估计。对于推测它们在原始条件下合成的可能性，其结构和组成已提供了主要线索。在图1-51的谱系树中，包括最初12个非极性的脂肪族氨基酸、以及较简单的极性侧链的几条亲缘关系分支。从这张图引出了使人感兴趣的、关于自然界选择蛋白质字母表的问题。

甘氨酸和丙氨酸这两种最简单的氨基酸是“自然的”代表。加入一些较高级的同类物，诸如天冬氨酸、亮氨酸和异亮氨酸等，显然就容易满足疏水相互作用的要求。这种特定的选择也许是从属于偶然性的，也许还由于与可利用的连接物的有辨别力的相互作用而得到加强。在极性侧链中，我们发现了一些明显

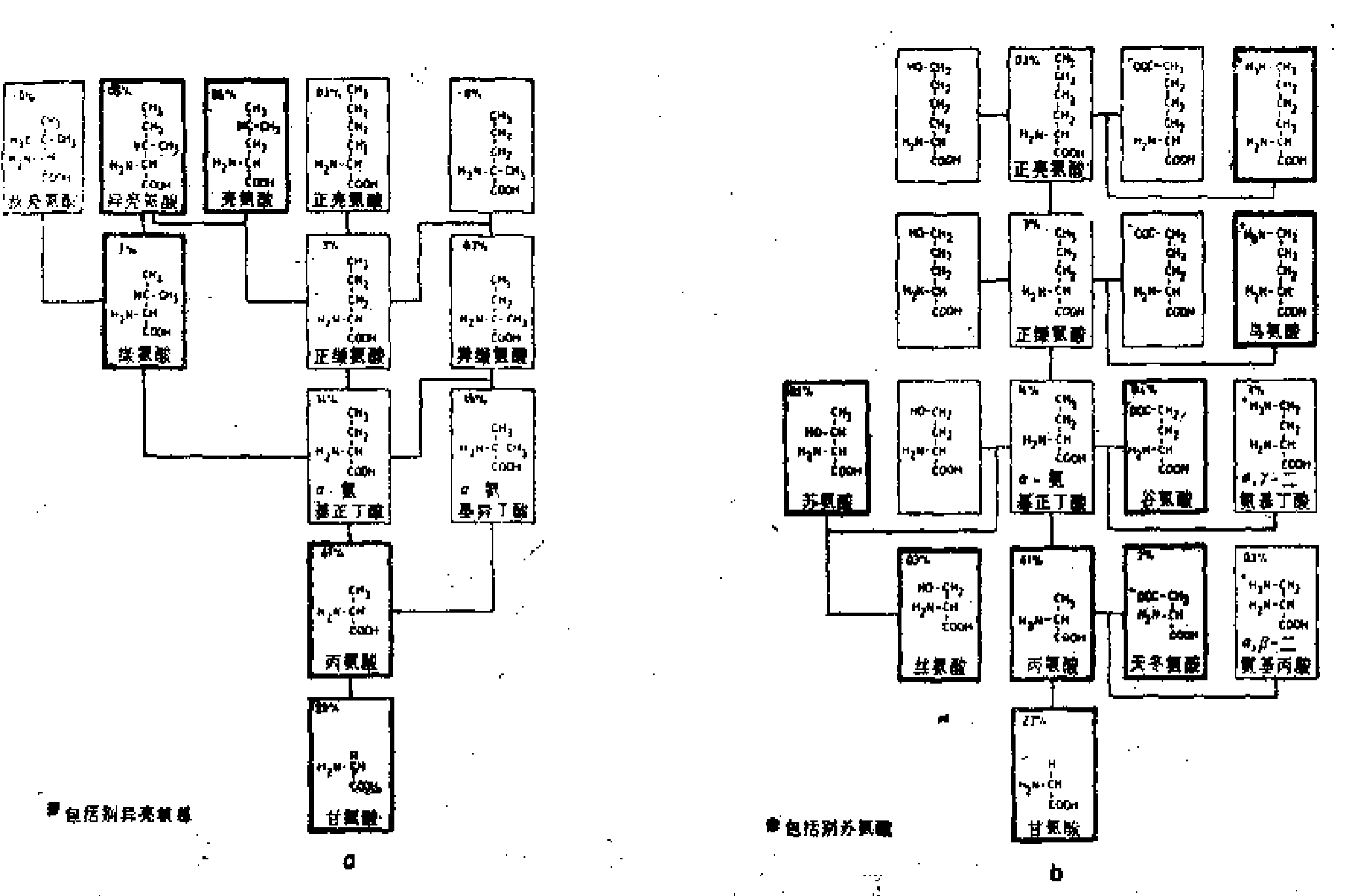


图1-51　最初的脂肪族氨基酸和一些最简单极性侧链分支的谱系树。在每一个方框中，左上角的数字指的是米勒在原始条件下的相对产率数据(即特定氨基酸的摩尔产率除以所有氨基酸的产率之和)。粗线条方框中是自然的氨基酸。

的脂肪族羧酸(天冬氨酸和谷氨酸)以及醇①(丝氨酸和苏氨酸)，而没有发现相应的胺②(α,β-二氨基丙酸和α,γ-二氨基丁酸)。只有下一列同类物(赖氨酸)出现在20种“自然的”氨基酸中，而中间物(鸟氨酸)仍然表现出它的痕迹。原因可能是，在活化的第二个氨基上，出现了内酰胺的生成或消失，从而中止了聚合。而且，此第2个氨基可能会导致多肽链出现分支(尽管对于羧基可能引出类似的论据)。在最初的功能多肽中，荷正电侧链可能并非必要。甚至在现在的海水中，对于引起与羧基的明显的复合作用来说，Mg^{2+}浓度是足够高了(约50mM)。在还原性条件下，海洋中甚至可能溶解了更多的2价离子(例如F^{2+})。那些附着在羧基上并仍然具有配位部位的金属离子，对于早期的蛋白质与(荷负电的)多肽之间的紧密相互作用是特别重要的。从这种观点来看，其中带有荷负电的配位体的侧链没有带有荷正电配位体的侧链那么必要。

图1-51中未列出的“自然的”氨基酸，它们带有复杂得多的侧链，因此在原始汤中以相当低的浓度存在。

这些根据结构和组成提出的推测，已从米勒和其他人所进行的模拟前生物合成氨基酸实验中得到了极好的确证。自然氨基酸(以及谱系树上的其他分支)的产率，大致合乎化学家的预期值(数据见图1-51)，这些实验还带来了许多有趣的详细信息。而且，这些结果与陨石分析所获得的数据相当一致，这种分析反映了星际空间的氨基酸的存在。表1-16列出了与我们的讨论有关的数据。

毫无疑问，原始汤中的丙氨酸和甘氨酸是非常丰富的。在米勒的实验中，这些氨基酸出现的频率，大约是其他“自然的”

①② 原文如此。——译者注

表 1-16 模拟前生物合成中和默奇森陨石中的自然氨基酸丰度

第一列指的是那些出现在蛋白质中的氨基酸。第二列的数据是米勒实验的典型结果。它们是在存在氮和水的情况下，对336毫摩尔甲烷进行放电而获得的。以碳为基准，氨基酸（包括那些蛋白质中不存在的氨基酸）的总产率为1.9%，相应的甘氨酸和丙氨酸的产率分别为0.26%和0.71%。表中所列的产率是指摩尔丰度。在不同的条件下仍获得类似数据，甘氨酸、丙氨酸、天冬氨酸和缬氨酸通常作为丰度最大的自然氨基酸出现。由奥罗等人和克温渥尔登等人所报道的陨石数据中，所有情况下 *D* 异构体都以接近50%出现。

化合物	产率（μM）	μg/g 陨石
甘氨酸	440	6
丙氨酸	790	3
天冬氨酸	34	2
缬氨酸	19.5	2
亮氨酸	11.3	
谷氨酸	7.7	3
丝氨酸	5.0	
异亮氨酸	4.8	
苏氨酸*	1.6	
脯氨酸	1.5	1

\# 包括别苏氨酸

氨基酸出现的频率的20倍。在自然氨基酸的丰度标度中的下两个位置，由于冬氨酸和缬氨酸占据，它们与亮氨酸、谷氨酸、丝氨酸、苏氨酸和脯氨酸之间有明显的空隙。

各种理由都支持作出如下假定：密码子对氨基酸的编码分配实际上遵从丰度标度。如果丙氨酸和甘氨酸是丰度最大的氨基酸，那么为什么一旦有可资利用的活化机制时，它们没有最先受到编码分配呢？最原始的多肽，包括受到指令的以及未受指令的，必定主要是丙氨酸-甘氨酸共聚物，只是偶然地被其他氨基酸其中也包括那些最后没有成为被编码分配的氨基酸所取

代。

引人注目的是，自然氨基酸丰度与最初4种密码子分配顺序的一致性。应该强调的是，我们选择密码子，全以那些与核酸结构相联系的论据为基础。不仅最初的4种GC框架密码子确实与丰度顺序完全吻合，而且4种外加的RNY密码子被分配给那些氨基酸，其中除了天冬酰胺以外，它们在米勒表中以大得可以看到的产率很好地体现出来。人们可能要追问，AA^{U}_{C}对天冬酰胺的分配究竟是一种原始的分配呢，还是如同与赖氨酸密码子密切相邻所提示的，本来是与任何一种较低级的二氨酸同类物相联系，这些同类物在米勒的表中其产率顺序与天冬氨酸(α,γ-二氨基丁酸)和谷氨酸(α，β-二氨基丙酸)①相反。不过，若无进一步的证据，这也许是一种过分深远的推测。

在这方面，核苷酸结合酶的蛋白质序列分析的一些新结果也是重要的，人们相信，这些酶在前细胞条件下已存在了3×10^9年以上。这些数据提示，存在着核苷酸结合表面的前体序列，除了异亮氨酸、赖氨酸和苏氨酸以外，其中还包括缬氨酸、天冬氨酸(和谷氨酸)、丙氨酸和甘氨酸(虽然这些数据涉及到的前细胞进化阶段，晚于本文中所讨论的阶段)。

3.5 早期翻译器中的超循环组织

关于早期密码和翻译器的任何进化模型，都必须提供一些条件，以使类tRNA连接物以及各种酶因子的基因前体(或信使)不仅仅可以共存，而且可以相干地生长并向优化功能进化。

① 原文如此。——译者注

在第一部分和第二部分中业已表明，这种自组织要求所有伙伴之间形成循环封闭的反应链，如果它们全部不能从结构上整合成为一个复制单元。在本节，我们必须表明，现实的密码模型是怎样与超循环组织相联系的，这样的系统是怎样进化的。但是，这里有一个明显的困难，即超循环是怎样起源的。用通俗的话来说，超循环起源的前提似乎是要富足地存在着它的各种组分。用更专业化的术语来说，超循环由于是高级的反应网络，所以它必须通过某种高级机制"集成"后才得以存在。

为了进行比较，考虑一个按照一级自催化速率定律生长的简单复制单元。在高能原料缓冲的溶液中，一个拷贝就足以开始增殖过程。已利用噬菌体RNA或它的非感染的变异体进行过那些实验。一条模板链足以在几分钟内产生出一大群同样的拷贝(参见第一部分)。

一个超循环决不可能以这种方式出现。如果没有充分大量的专一性催化的相应物，一个模板拷贝是不可能增殖的。反过来这些相应物又由模板编码，而若没有它们的翻译产物的帮助，则模板自己也是不可能增殖的。系统中的所有模板的生长都有赖于催化支持，如果模板不增殖，催化剂就不能生长。通过某种偶然涨落而集成的概率有多大？让我们假定，在一个试管中有1毫升样品溶液。大分子的受控扩散反应速率常数的数量级可以是$10^{8}M^{-1}s^{-1}$。因此，要在半天内开始模板的增殖，就要求在给定的催化反应作用物中至少要有10^{6}个同样的拷贝。从如此巨大的涨落巧合中形成若干种作用物的相关功能的机会是不存在的。当然，各种模板按照混合级项进行增殖是可能的；换言之：把一级(无酶的)自催化项(代表没有得到其他成员催化帮助的模板增殖)叠加在二级催化复制项上。只有在浓度升高到充分高的水平后，此种超循环连接才会变得有效。不过，此

系统不可能预先知道，在按照一级自催化增殖的许多可选择序列中，哪一个序列对于组织的后期阶段所要求的催化剂提供了有用的信息。

对这个问题只有一种解答：

超循环必须有某种前体，此前体以高自然丰度存在，超循环通过某种突变和选择机制从前体逐步地起源。

富GC序列分布所构成的拟种，的确可以成为这种前体。一个稳定拟种的所有成员都将生长，直到它们以高浓度存在。正如3.4节中表明的，某些富GC序列能够通过把氨基酸分配给一定的反密码子而开始翻译。在这个阶段，对于保持系统来说，翻译产物确实还不是必要的，所以翻译仍然能够看作是一种反复试验的博奕。不过，如果出现了其中一种翻译产物对于复制自己的信使提供了优势，那么这种信使可以成为拟种分布中占支配地位的代表。

一个RNA物种，如果正链和负链都可以作为两种互补的密码子(例如GGC和GCC)的连接物，那么它至多能够分配一张两种氨基酸的字母表。如果连接物序列的丰度充分大，那么也有某种有限的机会，共存的突变体分配 2 或 4 种密码子(包括天冬氨酸和缬氨酸的GAC和GUC)，同时又可能既利用正链又利用负链。所有这些，在拟种阶段仍可继续发生。

不过，只有不同的RNA物种借助于它们的翻译产物互相稳定时，这种系统才能进化。我们把关于分配细节的讨论推迟到3.6节中去，譬如，是不是一给定的RNA物种中的正链和负链都能够伴随着进化从而成为互补密码子的两个连接物呢，或者只是正链作为信使对耦合因子编码而负链作为连接物。在此我们研究的是，超循环组织怎样能够从某个拟种分布中逐步进化。

图1-52显示了如何能把这种过程形象化。假定有两种丰富

的拟种突变体，其正链和负链(大多数情况下)能够作为两种氨基酸对(例如，甘氨酸/丙氨酸和天冬氨酸/缬氨酸)的连接物，同时(大多数情况下)又可以被翻译成为由4类氨基酸构成的蛋白质；如果翻译产物提供了有利于它们的信使复制的某种催化功能，人们就很可能会遇到图1-52或者图1-53中所表示的一种情形。

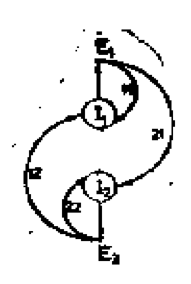


图 1-52　为它们自己的复制酶 E_1 和 E_2 编码的两个突变基因 I_1 和 I_2。可能会表现出自我增进和相互增进的等价耦合，这些增进作用是由于它们的近亲关系的缘故。类似的行为可在现在的RNA噬菌体复制酶中找到。

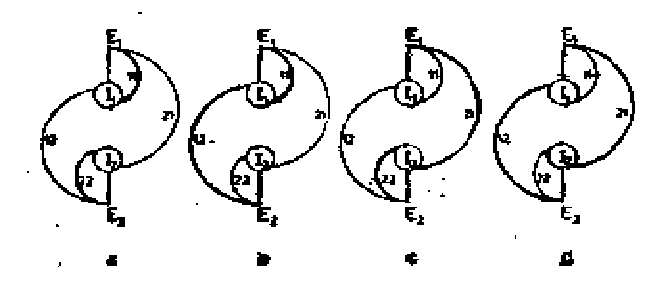


图 1-53　超循环的进化原理可由图1-52中显示的两个突变体之间耦合引起的4种可能情形来加以说明。粗线条表示优先的耦合(不论它多么小)。一个稳定的2元超循环要求优先考虑如 *d* 中所示意的相互增进。

两种信使，由于是密切相关的突变体，因此以密切相关的功能编码两种蛋白体。如果其中一个是一种专一的复制酶，则另一个也将是一种专一的复制酶，两者的功能是自我增进和相互增进的。不过，因为两种蛋白质不必同样好地识别两种序列，

它们也完全不必识别无关的序列，所以也可能有专一性。这些序列对于起始复制提供了某种特定的结合部位。E_1和E_2与I_1和I_2的4种可能的相互作用中的结合强度差异可以很小。由线条强度所表示的这些差异，无论怎么小，都将有惊人的后果，正如对相应的固定点图(表1-17)进行考察所表明的。我们可以区别4种情况：

表 1-17　图 1-52 中所示的 2 元超循环的固定点分析已运用简化的速率方程得到实现。

$$\dot{x}_i=\sum_{k=1,2}k_{ik}x_ix_k-\frac{x_i}{c}\sum_{l=1,2}\sum_{m=1,2}k_{lm}x_lx_m$$

$i=1,\ 2;\ x_1+x_2=c$

得出3个固定点及其本征值：

$\bar{x}_1=(c,\ 0);\ \omega^{(1)}=(k_{21}-k_{11})c$

$\bar{x}_2=(0,\ c);\ \omega^{(2)}=(k_{12}-k_{22})c$

$$\bar{x}_3=(k_{22}-k_{12},\ k_{11}-k_{21})\frac{c}{k_{11}-k_{21}+k_{22}-k_{12}}$$

$$\omega^{(3)}=\frac{(k_{11}-k_{21})(k_{22}-k_{12})}{k_{11}-k_{21}+k_{22}-k_{12}}c$$

可以区别出4种情况：

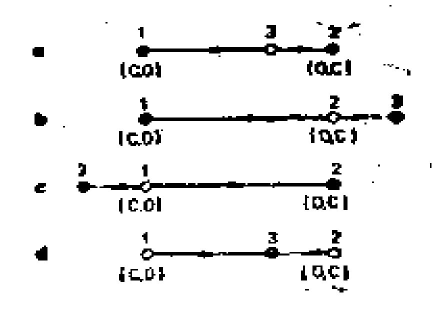
a. $k_{11}>k_{21};\ k_{22}>k_{12}$　　I_1和I_2竞争

b. $k_{11}>k_{21};\ k_{22}<k_{12}$　　选择I_1

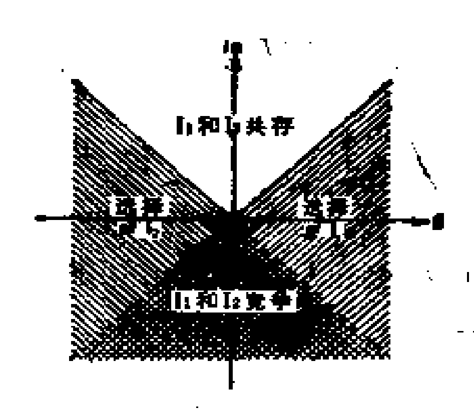
c. $k_{11}<k_{21};\ k_{22}>k_{12}$　　选择I_2

d. $k_{11}<k_{21};\ k_{22}<k_{12}$　　超循环稳定了I_1和I_2

这4种情况的固定点图是(参见第二部分)：



如果引入两个坐标：$\alpha=k_{12}+k_{21}-k_{11}-k_{22}$ 和 $\beta=k_{12}-k_{21}+k_{11}-k_{22}$；则可获得统一的表示。

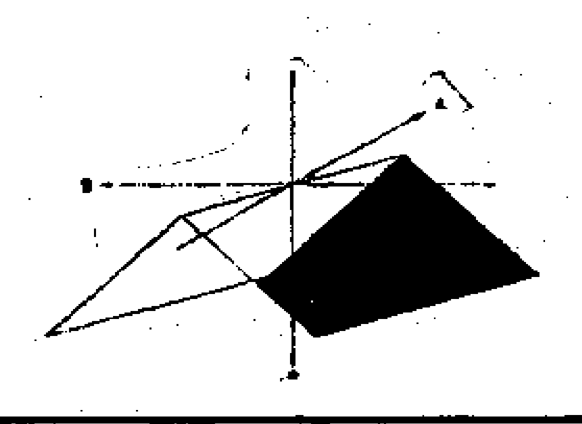


如果在此速率方程的二次项中加上一个线性的自催化项（得出 $\Gamma_i=k_ix_i+\sum_{j=1,2}k_{ij}x_ix_j$ 形式的生长函数），则两物种 I_1 和 I_2 的稳定的超循环共存区域处于如下坐标的3维参数空间的折叠面之上：

$$\alpha=k_{12}+k_{21}-k_{11}-k_{22}$$

$$\beta=k_{12}-k_{21}+k_{11}-k_{22}$$

$$\gamma=\frac{2}{c}(k_1-k_2)$$



---

(1) E_1有利于I_1而不是I_2，E_2有利于I_2 而不是I_1(图1–53*a*)

结果：I_1和 I_2 两者由它们相应的酶超循环地加强，从而导致强烈的竞争。即使两个竞争者是选择上等价的，也只有其中

一个竞争者能够生存下来。

(2) E_1有利于I_1而不是I_2，E_2也有利于I_1而不是I_2(图1-53*b*)

结果：I_1将赢得竞争，而I_2将消亡。

(3) E_2有利于I_2而不是I_1，E_1也有利于I_2而不是I_1(图1-53*c*)

结果：现在的获胜者是I_2，而I_1消亡。

(4) E_1有利于I_2而不是I_1，E_2有利于I_1而不是I_2(图1-53*d*)

结果：我们在此获得了I_1和I_2的相互的超循环稳定。

注意到小的差异足以引起上述行为是重要的。在此方面有趣的是发现当E_1和E_2对待I_1和I_2完全等价时所发生的情况。在此，我们是完全不偏袒任何一方的，而无论群体数x_i和y_i是多么不同。一个涨落突变结果就能使I_1和I_2消亡，因为在此不存在象第4种情况中那样的相互稳定作用。另一方面，这些涨落不能放大自已，并且，如果群体数充分大，则涨落突变实际上决不会发生。如上面所述，情况(1)在这方面是相当不同的。只有对于I_1、I_2、E_1和E_2的群体数正好相等，此系统才处于动力学上的平衡态。一个小的涨落可以扰动此平衡，并通过自我放大，从而不可避免地导致选择两种物种中的一种。这也适合于其中每一个信使只对其自己的复制酶提供帮助的任何系综（参见图1-47)。所有复制酶(利用其RNA识别部位)可能同时作为活化酶起作用，这种共同的翻译功能所形成的耦合，不足以促进某种共存。如同图1-45所示的系统中，将只有一种存活者，翻译功能也将相继被破坏。

表1-17中导出了形成超循环的准确标准。那些数字清楚地表示了坐标的稳定区，坐标是推广了的，其中包含速率参数。

因此，我们获得了超循环进化原理。只要发展反应耦合的方式，这种组织就能够从某种拟种分布中崛起。只有通过密切相关的突变体，才能一般地满足前体共存的前提。从而超循环

的出现要求先存在某种分子达尔文系统，但是超循环将导致相当新的结果。哪怕在速率参数中只有非常小的差异，此进化原理也是有效的。因此，此进化原理对于由突变体带来的小变化的反应是灵敏的。给定一种拟种分布，它已发展起了在各组分之中的相互作用；无论这些相互作用是多么微弱，一旦产生，超循环组织就将不可避免地出现。

超循环也将不可避免地通过突变的方式向更高的复杂性生长（图1-54，图1-55）。通过归纳，此进化原理能够加以推广，

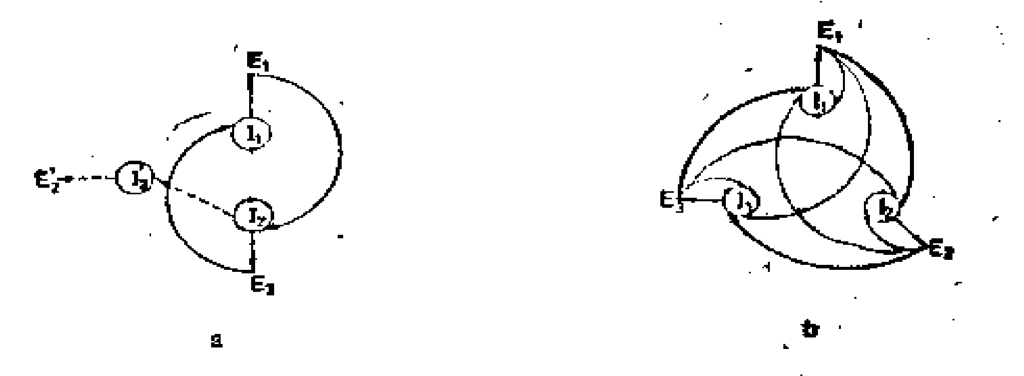


图 1-54　超循环进化原理的推广在此图中得到说明。耦合必须满足表1-17和表1-18中导出的标准　即相互增进必须强于自我增进(参见粗线条)。a.出现了I_2的突变体(I_2')。b.此突变体（现在为I_3）并入了此超循环中。

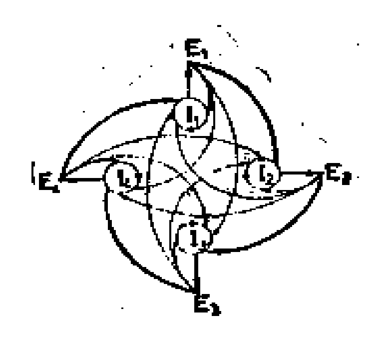


图 1-55　"现实的"4元超循环将4个信使I_1到I_4(是某种普通前体的突变体)，分配给编码具有相同功能的4个复制酶E_1到E_4，但专一性中稍有偏向。I_1到I_4的负链相伴地作为氨基酸连接物。

以适用于任何n元超循环。于是，某个突变体I'可以或是取代掉I，或是死亡，或是把此超循环扩大为$n+1$元的超循环（参见3.6.10）。正如表1-18中表明的，能够导出更一般的进化标准。

表1-18　图54示意的一个2元系统转变成为一个3元系统，解释了超循环进化原理的推广

一般的速率方程与表1-17中具有相同的形式。从一个稳定的2元超循环出发，引入了第3个成员I_3（例如，它作为I_2的一个突变体）将反复地把先前稳定的固定点转变成某个鞍点。如果假定了循环的对称，就会很清楚地看到这一点。在这些条件下，其记法可以简化为：

$k_{11}=k_{22}=k_{33}=k_D$；$k_{13}=k_{21}=k_{32}=k_+$；

$k_{12}=k_{23}=k_{31}=k_-$；

它产生出如下的速率系数矩阵

$$K=\begin{vmatrix} k_D & k_- & k_+ \\ k_+ & k_D & k_- \\ k_- & k_+ & k_D \end{vmatrix}$$

固定点和本征值则是：

隅角：$\bar{x}_1=(c,\ 0,\ 0)$；$\omega^{(1)}{}_1=(k_--k_D)c$，

$$\omega^{(1)}{}_2=(k_+-k_D)c,$$

$\bar{x}_2$、$\bar{x}_3$与此相类似

边：$\bar{x}_4=(0,\ k_D-k_-,\ k_D-k_+)\dfrac{c}{2k_D-k_+-k_-}$

$$\omega^{(4)}{}_1=\frac{k_-(k_D-k_-)+k_+(k_D-k_+)+k_+k_--k_D^2}{2k_D-k_+-k_-}c$$

$$\omega^{(4)}{}_2=\frac{(k_D-k_+)(k_D-k_-)}{2k_D-k_+-k_-}c$$

$\bar{x}_5$、$\bar{x}_6$与此类似

内域：$\bar{x}_7=\left(\dfrac{c}{3},\ \dfrac{c}{3},\ \dfrac{c}{3}\right)$

$$\omega^{(7)}{}_{1,2}=\{2k_D-k_+-k_-\pm i\sqrt{3}(k_+-k_-)\}\frac{c}{6}$$

下列4种情况具有特别意义：

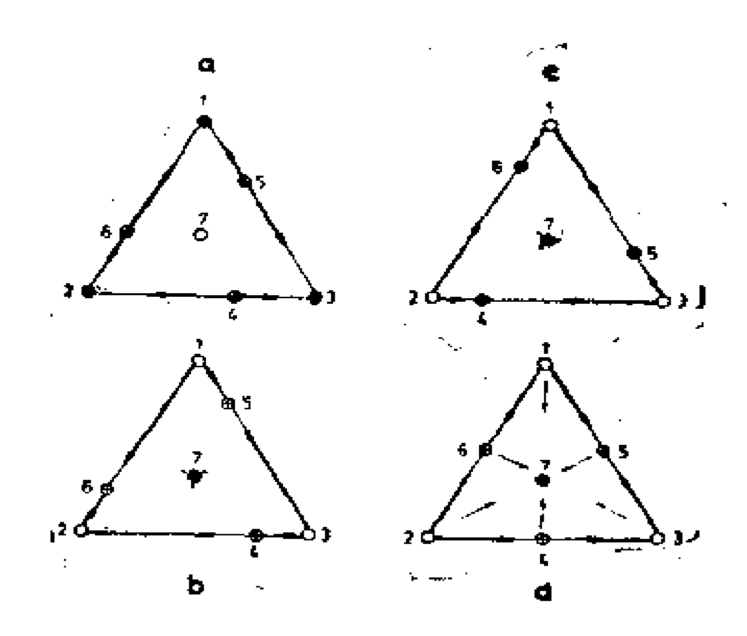
a. $k_D > k_+$, k_-

b. $k_+ > k_- > k_D$

c. $k_- > k_+ > k_D$

d. $k_+ = k_- > k_D$

它产生出如下的固定点图：



大的对角项($k_D \gg k_+$, k_-)导致竞争（图 *a*）。在相反的情形时，即K具有大的非对角元素时，这3种物种表现出协同行为。旋绕着此单形中心的螺旋收点的情形，是由两个常数 k_+ 和 k_- 中较大的一个所决定的。对于相等的常数 $k_+ = k_-$，观察不到旋转分量，所以中心固定点是一个焦点。

这张表中所处理的例子，很好说明了向更复杂的超循环进化。在未对速率常数作简化假定的情况下，此分析变得相当复杂。我们指的是包括推广到任意维数的更详细的阐述。

3.6 十个问题

涉及到极早期的分子原型以及它们遗留于现在细胞中的生物合成器中的痕迹。

3.6.1 一种RNA前体？

这个问题涉及到开始具有某种可复制功能的最初分子的复杂性。一条100个残基的核苷酸链相应的复杂性约为10^{60}种可选择的序列。如果以稳定性为基础，我们只把自己限制在(含有AU)GC共聚物，那么我们仍然留下了约10^{30}种可能的排列。为了取得一个或几个确定的序列，精确的自复制是必要前提。它将不可避免地导致达尔文行为，因而在一种确定的拟种中进行选择。被选择的产物只不过由某种最优的选择效率加以决定，但是它们的结构则依赖于它们的历史途径，这还受到较小的低聚核苷酸模式的自复制的强烈影响。

3.6.2 选择优势对于分子意味着什么？

选择价值被定义为结构稳定性与精确复制的效率的某种最优组合。它能够以与一定环境中的分子物理性质相联系的项来加以定量地表示。结构稳定、抵制水解和协同性质的发展要求延伸作用。小的低聚核苷酸不可能以任何稳定的方式折叠，因此易于水解。而且，对于忠实复制或者翻译，它们不可能提供充分大的粘接强度。另一方面，强度又受复制速率和复制精确度的限制。对于形成具有延伸长度的稳定拷贝，已表明富GC序列的性质是有利的。这些长度是否跟今天的tRNA相当，目前还不能肯定。在tRNA中已找到了序列同源物，这标志了内部区域的某种自复制。不过，这也许在密码子被分配之前就已发生了。翻译开始起作用，需要连接物和信使之间有强烈的相互作用，而太小的分子是不可能提供这一切的。

只要翻译产生出可复制的功能，选择价值就达到了一个新尺度。然而，对于任何给定的信使，它必须以忠实复制的结构稳定性和效率来加以表达。不过，这些性质在此也取决于翻译产物的性质(和浓度)。超循环组织所需要的特定耦合对于任何系统都是必需的，在这些系统中，翻译产物被看作用于选择，并由此适合于进化。这样的耦合具有催化性或保护性。

3.6.3 为什么是单突变基因而不是一个稳步生长基因组的超循环组织

对这个问题的回答基本上已在第一部分中给出。对于非常原始的翻译器，所需要的信息量相当于（或者甚至超过）现在的RNA噬菌体的信息量。只有在噬菌体专一酶复合体的帮助下，此种噬菌体基因组的信息才能够得到保持，信息的可利用性，是以由宿主细胞提供完整的翻译机构的有效性为基础的。如果我们接受了关于第 1 个问题和第 2 个问题的答案，则开始翻译所需要的信息必定来自共存于拟种分布中的几种突变体之间的协同作用，而不是只来自一条序列的延伸，因为对于它本来不存在选择压力。

几种共存的突变体的超循环稳定化，等价于通过基因复制的进化。本来，突变体是作为单链而不是作为共价键连接的复制体出现的。精确度的限制，将不允许这种长度延伸。而且，获得一条链中所需要的突变体组合的概率非常小。由100个G和C残基组成的序列有

100个单错突变体

4950个双错突变体

161700个三错突变体，等等，或者

$$N_k=\binom{100}{k}\text{个 } k \text{ 误差突变体}$$

包含了n个突变基因的链的条数，其中每条都与另一条有k个特定的位置不同(为定出一个函数这可能是必要的)，这样的链数总计为

$$\binom{N_k+n-1}{n}\approx\frac{N_k^n}{n!}\qquad(\text{对于 } n\ll N_k)$$

例如，对于$n=4,k=3$，共计有3×10^{19}种可选择的序列。即使在增殖基因中给定这些小偏离，在巨大的一条链中发现一个具备有利组合拷贝的机会，对于任何合理大小的群体而言，都几乎为零。不过，每一个此种包含了3种取代物的孤立突变基因，在任何宏观群体中都会大量地存在。

最后但不是最不重要的一点，作为翻译连接物的tRNA必定已作为分离的链而出现。统一的基因组的进化，也许在一开始就要求复杂的转录控制权。

另一方面，独立的RNA链可以在拟种分布中自然起源。所有的序列都是类似的，所以它们的翻译产物亦必定是类似的。只要一个翻译产物提供了耦合功能，则所有的翻译产物，由于它们的类似性，都将提供耦合功能。作为超循环组织所需要的循环耦合，从而也可以出现。我们也许甚至可以认为，超循环组织极其自然地与任何实际的原始翻译模型相联系。

在原核细胞中建立起来的现有基因组织，对它的早期结构是否提供某种线索？现在的基因肯定比早期的信使大得多。只要酶机构稳步地增加精确度允许基因的延伸以及复制，就会提供某种优势。于是就能获得更为高级的翻译产物，并可利用从共同前体传下来的有差异的酶向更为复杂的多酶机构进化。现在的细胞所利用的重组机构，在原始系统还不存在。因此，现

今原核基因组的结构可能是通过下列过程获得的：独立基因的延伸，它们的复制及一式三份的过程直到操纵子，以及最终它们的基因定位到*DNA*上。它能够利用更为先进的复制方式以致允许形成某种统一的基因组。现有操纵子的大小，正好相应于能够由高级的*RNA*复制酶所处理的大小(例如1000到10000个核苷酸)。

3.6.4 开始时*tRNA*是必要的吗？

这个问题可以换个说法：为什么没有小的低聚核苷酸连接物？

无信使的连接物没有多少意义。短的核苷酸序列不配作为信使。10肽已几乎相当于半个*tRNA*分子。而且，因为短的低聚核苷酸没有任何三级结构，所以它们可能是不稳定的。最简单的对称结构即由4个或5个碱基对稳定的一个环，需要15个之多的核苷酸。对某个氨基酸作无酶专门识别，这同时涉及到反密码子环和连接物的分端，而且只有更加伸展了的结构才有可能作出识别。对于两个邻近的连接物之间的相互作用亦是如此，对于信使-肽酰*tRNA*复合体的稳定化也是必要的，或者对于构象变化(例如从*HF*变为*fh*)也是必要的。这种构象变化可以促进生长着的肽链沿信使的输运过程。马斯等人最近报道了在*tRNA*的反密码子环中的这种构象变化，他们通过观察*Y*碱基的荧光变化而记录了这种现象。这种效应看起来在反密码子环片段(即具有反密码子环序列的10肽)中是没有的。所有这一切提示，不是充分长的RNA序列，就不配具有连接物功能。

从而我们可以问道：连接物与信使的区别是什么？它们都要求类似的最小长度。它们都必须是特定的折叠结构，通过这

种结构它们可以重复地被耦合因子识别。

因为每一个tRNA和每一个信使要求一个耦合因子,例如要求一个复制酶从而有利于它们的选择稳定化，所以 RNA 序列的双重功能是必不可少的。因此一条给定的 RNA 序列的正链和负链便可以联合地被利用作为信使和连接物。

3.6.5 现在的*tRNA*对其起源提供了线索吗?

结构上的类似性，也许是对一个共同目标适应的结果，或者也许是标志了一种共同的前体。现在的 tRNA 在其结构上表现出许多一致之处。我们能够从这些类似中作出某种相应的关于前体的推论吗?按照由朱克斯进行的分析，这个问题可以用谨慎的"是"来回答。为什么人们必须谨慎呢，这可用一个例子来说明。由迄今所研究的所有原核的和真核的*tRNA*,展示了一个共同特征,这就是所谓的*T*环中的序列*TψCG*,这是核糖体控制的一个共同识别部位。最近对甲醇基因细菌的研究揭示出，这些被看作"在细菌株系迄今所遇到的最古老的趋异物"的微生物，缺乏这种*RNA*的共同特征，而在一群中包含了一条序列*ψψCG*，在另一个群中含有一条序列*UψCG*。尽管这个发现是无可怀疑的，但毋宁说这个发现强调了这类微生物与其他原核生物密切的进化关系；它明确地显示了，所有各类微生物可以一致地取得共同的特征。对于那些由共同机构产生的诸如核糖体(它是所有蛋白质分子合成的地方)那样的分子，尤其是这样。

图 1-56 显示了大肠杆菌中的 4 个 *tRNA* 序列的线性排列，这被我们看作早期密码子连接物的现有代表。不幸的是，无法利用连接到密码子GCC上的丙氨酸专一的*tRNA*序列。如果我们把这个具有反密码子$^{5'}$*UGC* 的物种，与它的对应物进行比较这

```
     5    10   15   20     25   30   35   40
GCGGGAAUAGCUCAGDDGGD AGAGCACGACCUUGCCAAGG
GGGGGCAUAGCUCAGCDGGG AGAGCGCCUGCUUUGCACGC
GGAGCGGSAGUUCAGDCGGDDAGAAUACCUGCCUGUCACGC
GCGUCCGSAGCUCAGDDGGDDAGAGCACCACCUUGACAUGG

  45   50   55   60   65   70   75
UCGGGGUCGCGAGTQCGAGUCUCGUUUCCCGCUCCA
UCGAGGUCUGCGGTQCGAUCCCGCGCGCUCCCACCA
AGGGGUCGCGGGTQCGAAUCCCGGQCCGUUCCGCCA
UCGGGGUCGGUGGTQCGAGUCCACUCGGACGCACCA
```

图 1-56 甘氨酸、丙氨酸、天冬氨酸和缬氨酸的 tRNA 序列的线性排列 不幸的是，无法利用涉及到密码子 GCC（对于丙氨酸）的序列。甘氨酸 tRNA 和丙氨酸 tRNA 之间的对应物被假定更接近于涉及到反密码子 GCC 的正确序列（这是由分别涉及反密码子 *UGC 和 *UAC 的丙氨酸和缬氨酸的两种序列之间的类似性所提示的）。这些序列表明，碱基配对的区域主要由GC所构成；这些序列还表明，接近的对应物标志了甘氨酸/丙氨酸和天冬氨酸/缬氨酸之间的亲缘关系（参见，天冬氨酸和缬氨酸处于位置 8 的 S 而不是甘氨酸和丙氨酸的 U，或者在天冬氨酸和缬氨酸的位置20和21 之间的嵌入物 D）。A＝三磷酸腺苷，*A＝2MA＝N(2) 甲基三磷酸腺苷，C＝胞苷，D＝5,6-二羟基尿苷，G＝鸟苷，*G＝7MG＝N(7)-甲基鸟苷，Q＝ψ＝假尿苷，S＝硫尿苷，T＝核糖鸟苷，U＝尿苷，*U＝5AU＝5-羧乙酰尿苷。

种对应物用于缬氨酸，并且有反密码子 ^{5A}UAC，那么我们就会观察到，与线性排列中所列出的序列（与 54 相对的57是等同的位置）比较而言，它与后者更为一致。因此，具有反密码子GGC 的正确的丙氨酸 tRNA，与 44 位置比较而言，它与表列的甘氨酸 tRNA 更为一致。除了这种“实实在在的缺陷”之外，这些数据还揭示了：

1．所有的典型在一半以上位置是一致的（包括“错误的”丙氨酸有33个，或者对于甘氨酸、天冬氨酸和缬氨酸为41个），

2．亚基亮氨酸/丙氨酸由几种特征与亚基天冬 氨 酸/缬氨酸相区别（在位置 8 是硫尿苷“S”而不是U，在20到21位置之间

嵌入5,6-二羟基U),

3．所有的典型，尤其在其碱基配对区域，残基G和C都明显多于A和U(或它们的衍生物)。

进一步与其他tRNA 序列进行比较表明,尽管这些特征肯定在大多数tRNA 中是共同的，但是它们对于这种基因尤为显著。特别是，密切的对应关系就象同一有机体中的同一氨基酸的不同连接物一样。

有一个发现特别能说明问题。如果我们把这两种连接物的序列与互补的反密码子(例如天冬氨酸和缬氨酸)进行比较，则在tRNA的两条正链的一致性,比一条正链(从3′读到5′)和另一条负链(从5′读到3′)之间的一致性要明显得多。实际上，如果我们以这种方式比较相同tRNA的正链和负链,则有更好的一致性。这些一致性只不过表达了tRNA显著的内部对称性,其中反密码子几乎正好置于序列的中间，并由此允许形成对称的二维模式。我们可把这种性质当作 tRNA 早期作为一种独立的复制单元出现的标志。要求正链和负链采取某种类似的模式，如果这些模式代表了独立的复制单元而不是作为结构上整合到某条大的基因组序列中去(如同它们现在的样子)，这才是重要的。我们发现噬菌体RNA或它们的变异体有一种类似的效应，它们必须作为单复制单元而增殖。

另一方面,tRNA对共同机构的适应必定已引起了共同偏离起源时所要求的对称性。同一种tRNA的正链和负链的镜象,比具有互补的反密码子的tRNA的正链和负链的镜象,仍然显示了更为对称的相似性；这一事实提示，两种 tRNA 作为相同的而不是互补的两条链的突变体而进化。于是我们可以得出结论：密码子GGC(甘氨酸)、GCC(丙氨酸)、GAC(天冬氨酸)和 GUC(缬氨酸)的现有连接物，是从某一共同祖先的单错突变体的一

种拟种衍生出来的。不过，原始的对称性不足以(为什么它是这样?)允许从某一给定的 RNA 正链和负链两者衍生出连接物功能。

3.6.6 如何能够产生无逗点信使模式?

最初的信使必定是与最初的连接物(或它们的互补链)一致的。在连接物和信使功能背后的确有某种结构的一致性。不管在信使序列中出现何种密码子模式，它在连接物上必须有它的互补代表。在原始系统中，通过利用两类分子的某种共同结构模式，容易满足上述要求，这样，连接物便是最初信使的负链(如果我们定义正链总是与一条信息相联系)，一定的结构对称性亦允许正链和负链两者由相同的复制酶所识别。

最初伸展的*RNA*分子富含*G*和*C*，这是以结构稳定性和复制精确度标准为基础的某种选择结果。带有诸如*GGC*/*GCC* 那样的共同密码子模式的分子，需要引物指令(例如，由催化剂或暴露的 *RNA* 环提供)以及随后的内复制。这将不可避免地导致包含至少带有内部互补性的两种密码子模式，例如 5′*GGC*3′ 和 3′*CCG*5′的结构。

在噬菌体复制酶从头合成并放大RNA序列中，有一个很好的关于内模式复制效率的例子。如果Q_β复制酶从任何模板上大量脱落，那么它开始“编织”它自己的引物，然后有选择地复制和放大它们，直至最后出现了均一的宏观RNA序列群体，这种序列长度为几百个核苷酸。在不同的环境条件下，获得了不同的(但是均一的)序列分布。西格尔曼和米尔斯及其同事已经按序列排好了一些这种“中变体”，它们都含有Q_β复制酶的专一识别部位。进一步的实验已经揭示了这种从头合成的机制，实验表

明，对应于由此种酶识别的序列的小片段是作为引物而被制造的，然后在内部被复制和有选择地被放大。特别是早期的研究已经表明，序列CCC(C)和UUCG能够被此种酶识别。UUGG对应于所有tRNA共同的序列TψCG，并被看作是专一地与核糖延伸因子 EFTu 相互作用，这在Q_β复制酶复合体中是作为亚单元起作用的。具有只由上述两种低聚核苷酸及其互补片段GGG(G)和CGAA构成的序列的中变体的线性排列表明，在3/4以上的这些位置都是一致的，这标志了引物序列(图1-57)的内复制效率。

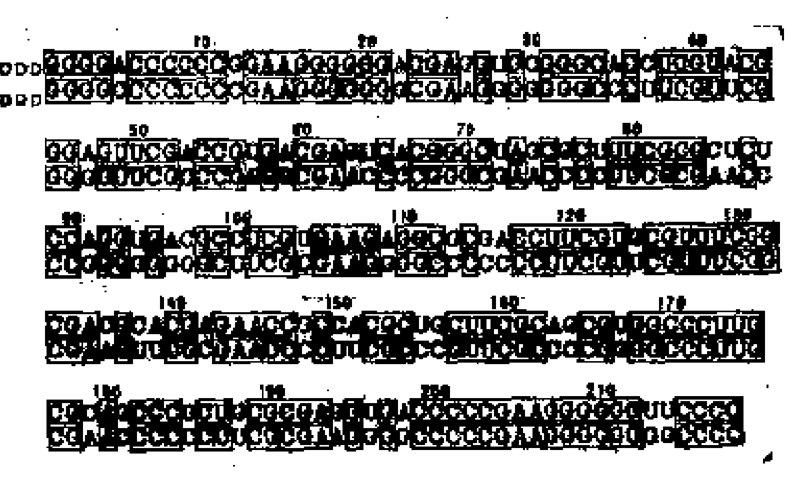


图 1-57 Qβ中变体序列的直线排列（由西格尔曼等人确定的），这是由CCC(C)条块和 UUCG 条块以及它们的互补物[GGG(G)和 CGAA]构成的人造序列。在218个位置中有169处一致提示，中变体是由Qβ复制酶从头合成的产物，它具有CCC(C)和 UUCG(EFFu)识别部位。从头合成的动力学表明，在酶的识别部位形成了四聚体，继之有某种具有偶尔取代的内部自复制。特殊的中变体通常在所有出现的序列竞争中取胜，因此好象是最有效的模板。此过程表明，在原始的复制机制中如何能够产生均一的模式。

类似地，我们可以考虑存在均一模式产生的原始机制。如果在许多可能的模式之中，出现了5′GGC/5′GCC以及可能还

有5′GAC/5′GUC，那么，那些信使模式可能已按照克里克等人提出的机制开始了某种可复制的翻译，并能够在它们的可复制翻译产物的帮助下进行选择放大。

3.6.7 最初的功能活性蛋白质看上去象什么？

最简单的蛋白质可以是某种均一的多肽例如聚甘氨酸。它可能提供任何催化活性吗？这是一个能够并应该由实验来回答的问题。运用混合的序列，包括充分多的比方说15到30个残基，可以形成具有一个活性中心的β片层结构，其中端位羧基置于一个靠近末端氨基的确定位置上（图 1-58）。相邻距离随链长而变化，因为β结构涉及到在两条反平行链中间的扭曲。末端氨基的 pK 值大约为 8，所以此催化部位至少含有一个有效的质子供体-受体系统。选用甘氨酸-丙氨酸残基对于形成β结构是非常有利的。不过，对于只由甘氨酸和丙氨酸构成的链，溶解性是一个严重问题，因为这将把它们只限制在界面。

周和法斯曼研究了β片层的折叠，他们为了阐明在链倒转

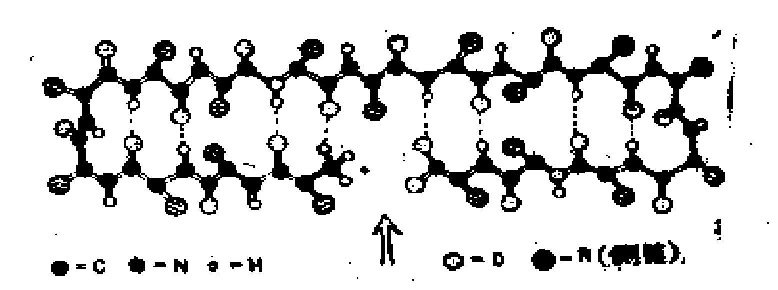


图 1-58 一个简单的酶前体大约由 15 到 25 个氨基酸（需要 45 到 75 个核苷酸的信使）的β折叠结构来表示。活性部位中包括十分有效的质子供体(pK～8)末端氨基，作为质子受体的末端羧基，以及一条催化活性侧链（例如天冬氨酸或丝氨酸）。可以设计出多种替换物，但其中只有一部分有可以产生出某种有效的活性部位正确的扭曲链的螺距。

区中的459β-转角而分析了29种蛋白质的X射线数据。在弯曲处的所有4个位置中的3个具有最高β-转角势的残基包括甘氨酸和天冬氨酸，而在β-转角区之后主要是疏水的残基。

末端基团有一定的空间排列，是催化效率的一个重要前提。对于稳定某种可复制的折叠，有必要利用两类或更多类的氨基酸。早已知道β片层结构是蛋白质的重要构件。按照莱维特的看法，它们可以以非常一般的方式被利用，以稳定蛋白质的活性构象。

高丰度的甘氨酸和丙氨酸也许在根本上已决定了最初蛋白质的外貌。但是，极性侧链对于较长序列的溶解性是必不可少的。四类氨基酸当然能提供多得多的灵活性。如果天冬氨酸和缬氨酸是随后的两个候补者的话，则也许已形成了球状结构，球状结构由缬氨酸和丙氨酸侧链的疏水相互作用而得到稳定，由天冬氨酸的羧基侧链而可以溶解。这种残基为2价金属离子参加下形成的专一催化部位提供了许多可能性。

估计如此巨大数目的可能性，对我们的想象是一种考验。为了检验各种结构在识别RNA序列和它们的结构特征时的效率，一些实验正在进行之中。由核糖核酸所获得的结果，鼓舞着人们去寻找某种能够专一识别RNA序列的“最小结构”。

3.6.8 合成酶对于开始是必要的吗?

现在的tRNA 3维结构中(参见第一部分图1-14)，反密码子环被固定在离氨酰部位相当远的地方。这种结构适应于现在的tRNA分子的功能需要，这是由核糖体机制和由合成酶结构所带来的。另一方面，众所周知，tRNA能够经受大大地改变其形状和大小的构象变化。里格勒及其同事通过荧光方法研究了构

象的寿命以及旋转弛豫时间，并作出了存在着至少 3 种不同的、迅速相互转化的构象状态的结论。奥尔森等人也获得了类似的结果，他们利用的是激光散射技术。不同构象状态的群体强烈地依赖于镁离子的浓度。重要的是必须再注意到，在相应于海水浓度的条件下(Mg^{2+}，约 50mM)，存在着一种构象体，由晶体学研究得知，它在形状上与L形式相异，即它更呈圆柱形。

强调这一点是因为它对所提出的这个问题十分有关。早期的酶只由数目十分有限的氨基酸残基构成，因此不可能具有非常庞大的球状结构。为了保护一种氨基酸对一种密码子的特殊分配，要么是必须得到象现在的氨酰合成酶那样高级的酶，要么是 tRNA 结构必须允许氨酰部位和反密码子部位比 L 形式有紧密得多的接触，以对两种部位进行同时的验证。否则，在早期阶段的高突变率，很快就又把这两种部位之间的任何独特的偶尔联系破坏掉了。另一方面，构象转变仍然是需要的，原因在于形成肽键的机制(参见图1-48)要求信使与生长的肽链明确分离开。所引的上述数据反映了这样的可能性。另一方面，如果很可能产生出类似于图1-49中显示的C 模式结构，那么甚至不借助于酶就已经得到了最初的氨基酸分配。这样的 tRNA 结构对于专一识别肯定提供了充分的敏感性。业已注意到，来自3′端的第 4 个碱基(即3′ACC后的碱基)无论如何总是与反密码子相联系的。最初期望的所有 tRNA 有一特殊的相互关系最终并没有实现。不过，这种相互关系可能由 *tRNA* 在早期专一识别氨基酸中起着关键的作用。考虑到来自大肠杆菌和T_4噬菌体的数据，在3′ACC 之后位置上的核苷酸是：关于甘氨酸的*U*、丙氨酸的*A*、天冬氨酸的*G*和缬氨酸的 *A*。它对于确保由有充分识别力的部位进行独特的分配的早期连接物肯定是重要的。这种性质也许已在进化的后期阶段中部分丢失了。这当然是一种推

测，有待实验证实。

结论是：在非常早的阶段时合成酶可能并非必要，但是最终可以证明*tRNA*是自然界从核酸制备酶的一次不成功的尝试。更为有效的识别力可能从耦合因子进化而来，并注定要专一识别类*tRNA*结构。

3.6.9 谁是最初的酶？

如果对于开创翻译，合成酶并不是真正必要的（这是一个重要的"如果"），那么留给我们的耦合因子很可能是复制酶，成为翻译相干进化绝对的基本要求。通过这种功能，在翻译产物中出现的某种选择优势，能够非常有效地反馈到信使。因此，专一的复制酶（它们都属于一类相似的蛋白质分子）不仅提供了超循环耦合的前提，而且对于蛋白质的进化也是非常重要的，原因在于只有它们才能告诉信使：什么是表现型优势，如何在基因型水平上为之进行选择，也就是加强合成特定的信使。正如在下一段将看见的，这种在基因型和表现型水平之间的耦合，如果与空间分离或分隔间相结合，则效果最好。

当然，随之我们还必须对各种翻译功能寻找催化支持。如果复制已与类tRNA信使（包括正链和负链两者）建立起明确的关系，它们的识别性质便可以很好地被用作合成酶和翻译酶（即，前核糖体的）功能。换言之，某种复制酶的基因复制物可以正好是某种合成酶信使的前体，以及诸如EF Tu那样的某种翻译因子的前体，由于复制酶和转移功能的化学作用是非常类似的，并且在现在的系统中看来是受到类似残基的影响，所以情况更是这样。

逐渐趋异的双重功能，可能曾是复制和翻译机制的某种非

常早期的要求，这正如要求基因复制是进化后期阶段的主要媒介物一样。

现在的细胞器中明显地留下了那些双重功能的痕迹，病毒也已利用它们作为在宿主细胞中的后生物进化。噬菌体Q_β的基因组只对其复制酶的一个亚单元进行编码，但是却利用了宿主细胞的3个以上的因子，这些已被鉴定出是核糖体蛋白S_1和延伸因子EF Tu和EF Ts。

比布里切尔研究了这些因子的性质，并发现它们利用了可识别tRNA分子这一性质，并同时包括在核糖体控制的数种功能之中。他证明，由噬菌体基因组编码的$Q\beta$复制酶的β因子，在大肠杆菌细胞中也有其前体；而且情况看来的确如此。运用免疫学技术，他还能够证明，一种含有EF Tu和EF Ts的蛋白质，其行为象未感染的大肠杆菌中的$Q\beta$复制酶的某个前体，并且看来包含大肠杆菌的某种未特别指出的RNA合成功能中。另外，对于合成酶也可以找到类似的对应物。看来某种功能一旦发展起诸如识别某些类型的RNA的能力，则自然界就会在别的任何需要它的地方利用这种功能（例如专一复制、核糖体输运和控制、氨基酸的激活）。

在某些方面，RNA噬菌体的形成酷似早期RNA信使的进化。噬菌体除了不能利用它们自己的基因组的专一识别能力（即通过专一复制的耦合）以外，它们尽可能多地利用宿主细胞的功能。不同的噬菌体（例如$Q\beta$、Ms2、R17）遗传不同的识别因子，尽管它们都是从宿主细胞中的某个共同祖先衍生下来的。在第一部分中它还表明，RNA噬菌体感染的最初阶段相当于一个简单的超循环放大过程。

3.6.10 为什么细胞最终有统一的基因组?

超循环通过信使系统的功能整合为扩大信息量提供了优势，超循环中的单个复制单元的长度是有限制的，这归因于有限的复制精确度。信息量增加从而可以建立起可再生的复制和翻译器，借助于它们，翻译产物能够向更高的效率进化。这将得到更高的精确度，反过来又将使每一复制单元的信息量增加，从而再增进了酶的品质。同时，正如3.5节表明的，超循环本身通过整合起更多的分化突变基因，而向更高的复杂性进化。

信息量的增加不仅将产生出更好的酶，它还可以允许一个复制单元继承一个以上的酶的信息，从而可从所列出的较早的进化约束条件中去掉双重功能。也就是说，根据它们翻译产物的特殊功能需要，复制的信使可以独立地发展。这也许是操纵子结构的起源。操纵子具有一种控制机制，它同时控制几个结构基因的复制，从而复制酶也许已经进化成为带有诱导或阻遏的专一控制因子的普通聚合酶。

认识到了功能耦合的优点以后，功能耦合似乎是开创任何翻译所要求的，我们应该问一问，为什么功能耦合最终又被全体基因完整的结构整合所取代了呢？哪怕是已知最原始细胞的基因组也是作为结构的单元。因此，超循环组织的限度在哪里？对它能够做出什么改进？

在一个被功能连接控制的系统中，我们必须区分两类突变。其中一类突变主要是改变信使自身的表现型性质，从而在某种专一复制酶或控制因子方面，变更了它的目标功能。这些突变在早期的进化阶段特别重要，原因在于*RNA*结构的表现型性质的重要作用。那些目标突变将立刻在选择上变得有效，有利的

突变被固定下来，而不利的突变则将会被消灭掉。

第2类突变，它们对于目标功能可以是中性的，也可以不是中性的，它们涉及到翻译产物中的表现型变化。信使的专一性愈强，突变愈可专一地变更翻译产物的功能。

突变体是否专一地得到选择的偏爱，仅仅取决于目标功能，而不论翻译产物是在有利的还是在不利的意义上被变更了，也不论它是否保持为中性。在前细胞进化的后期，最为普通的突变结果，将是翻译产物中的某些表现型变化，这种翻译产物是与某种未受影响的目标功能耦合的。于是突变体可以进一步增殖，但是如果证明它的翻译产物是不利的，它不会从以前的野生型中而被特别选择出来，此系统也不会抵制突变体进行选择。真正应该达到的，是将系统看成为一个整体而进行评价。由信使系统的空间分离，由小生境或者甚至更有效地由分隔间可以做到这一点。在一给定的分隔间中的信使，运用它的增生效率，能够使得它自己的翻译产物在环境中丰富起来，并与其他分隔间竞争。在有限的程度上，只通过空间分离这也是可能的。不过，一个分隔间若没有超循环组织便全然不会起作用。在分隔间的有限生活空间里，所有信使之间的强化的竞争，将破坏任何协同功能。

一个分隔间通过把它自己的再生产与它的总基因量的再复制联系起来，就能更有效地增殖。当然，这需要相当复杂的控制机制，把所有基因整合成为一个巨大的复制单元可促进这一点。这种整个分隔间的个体化，要求复制机构有高精确度。在第一部分中，我们把各种生命阶段的信息量与它们相应的（和观察到的）复制精确度（参见表1-4）进行了比较。

分隔间个体化很可能与从RNA基因或操纵子转变为DNA基因组相联系，原因在于只有DNA复制的机制能够充分保证高

精确度。个体化的新单元是整合的原细胞。耦合的结构和功能组织代替并改进了基因的先前的功能组织和基因产物。进一步研究环状排列的遗传图，还可以揭露一些结构组织起源的遗迹，哪怕重组的渐成效应也许已将这种痕迹掩盖了很多。

作为统一化作用和个体化作用的结果，细胞的(无性)增殖净增长服从一级自催化定律(在没有抑制效应时)。这种系统的达尔文性质允许选择进化以及大量各种分化物种共存。整合的细胞单元结果变得比更保守的超循环组织形式要优越。

另一方面，多细胞组织随后的进化又可能已利用了作为细胞新的亚单元超循环组织（非线性反应网络）的类似物或替代物，从而已在某些方面类似于分子自组织过程。

3.7 现实的边界条件

讨论"现实的超循环"，必须联系现实的边界条件，否则就是不完全的。我们将就此扼要谈一谈，这倒不是因为我们无视它们在进化的历史过程中的重要性，生命在我们的行星上出现毕竟是一件历史事件，只是因为我们意识到，我们没有多少东西可说。虽然生命的早期阶段由于进化的相干性，总会在现代有机体中留下某些痕迹，当然相应的早期环境的遗迹是不复存在了。在我们的讨论中，我们可能已把一些偏见加在模拟原始的、无模板的蛋白质合成实验上，这些蛋白质合成实验是由福克斯和其他人所实现的(参见多斯和劳赫富斯的评述)。我们的研究正是要理解早期的组织形式，它们允许生物合成机构能自复制、选择和进化适应，就象我们今天在活细胞中所遇见的情况。对于这种适应性的自组织，蛋白质并不具有这类适应性自组织的

基本物理前提，至少不象核酸那样明显。另一方面，它们的确继承了巨大的功能容量，在此它们大大优于核酸。因为在原始条件下形成蛋白质要容易得多，所以存在大量的各种催化物质必定成为某种基本的环境性质。这个领域中的研究已清楚地指出，在原始条件下可以出现相当有效的蛋白质催化剂。在这方面，特别值得对界面进行研究。如果它们由催化活性物质所覆盖，则它们可能曾作为最有利的原始合成部位。把分子运动限制在某个平面上，大大地增加了碰撞效率，在涉及到高级反应序列时尤其如此。

翁萨格业已强调，在原始条件下，海洋必定覆盖着聚积起来的亲水和疏水材料层。这些复合层必定为原始的制备化学提供了有利的条件。从界面所提供的明显优点来看，我们已经考察了在相应的环境边界条件下的超循环的性质。

作为一个简单的模型，我们考虑图1-59中示意的系统。聚合物合成被限制于仅在表面层($r=0$)，在此有一定的结合模板和酶的能力。除了我们必须清楚地说明扩散以外，动力学方程

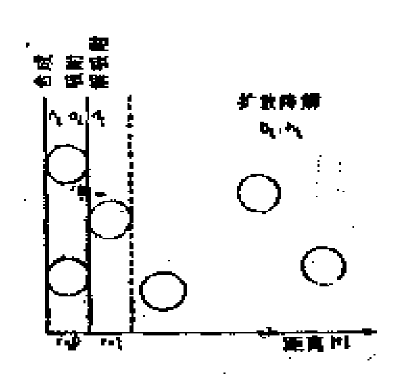


图 1-59　包括超循环耦合的非均匀反应模型示意图。
区分 3 个空间区域：$r=0$连接在界面上，
$r=1$ 界面的过渡层，$r>1$ 液相。
往返于界面的扩散被叠加在化学反应上，这些化学反应按照超循环模式进行。

与均匀溶液中的动力学方程相类似。我们辨认出与复制分子和酶的表面浓度有关的生长函数。假定表面上的扩散很快，没有确定的速率。大分子的吸附和解吸附被处理为表面层($r=0$)和接近表面的溶液层($0<r\leqslant1$)之间的交换反应。分解可以出现在界面和(或者只在)溶液中。最后，往返于界面的输运由扩散项来表示。

取决于所假定的合成机制，有必要考虑模板和酶的独立结合部位。我们运用这种模型去获得一些关于带翻译超循环的行为(参见2.5)的线索。按照有关文献所描述的方法，已对几组速率参数进行了数值积分。可以区分出3种特征性结果，其中两种完全类似于均匀溶液中的超循环行为：

1．在非常低的聚核苷酸和多肽浓度或大的K_i值时[见第二部分的(73)、(75)和(79)式]，聚合物的表面密度并不趋近稳态值，而是单调地或者阻尼振荡地减少。结果是大分子过一些时间后便消亡了(图1-60)。

2．在总浓度的一定阈值之上，我们在$n<4$的系统中发现了极限环行为。此种情形类似于在均相溶液中的低浓度极限(图1-61)。

3．在浓度充分高时，我们最终获得了某个定态：

$$\lim_{t\to\infty}\frac{\partial x_i}{\partial t}=0,\quad \lim_{t\to\infty}\frac{\partial y_i}{\partial t}=0$$

以及$\bar{x}_i>0$，$\bar{y}_i>0$，$i=1, 2, \ldots, n$ (图1-62)，x_i和y_i分别是酶和信使的浓度，$\bar{x}_i$和$\bar{y}_i$是它们最终的定态值，t是时间。

在较低维($n\leqslant4$)的系统中，只观察到类型1和类型3的行为。

这些模型计算，可以由运用随机计算机模拟技术研究一些密切有关的问题来补充。结果又一次表明，在界面上和均相溶

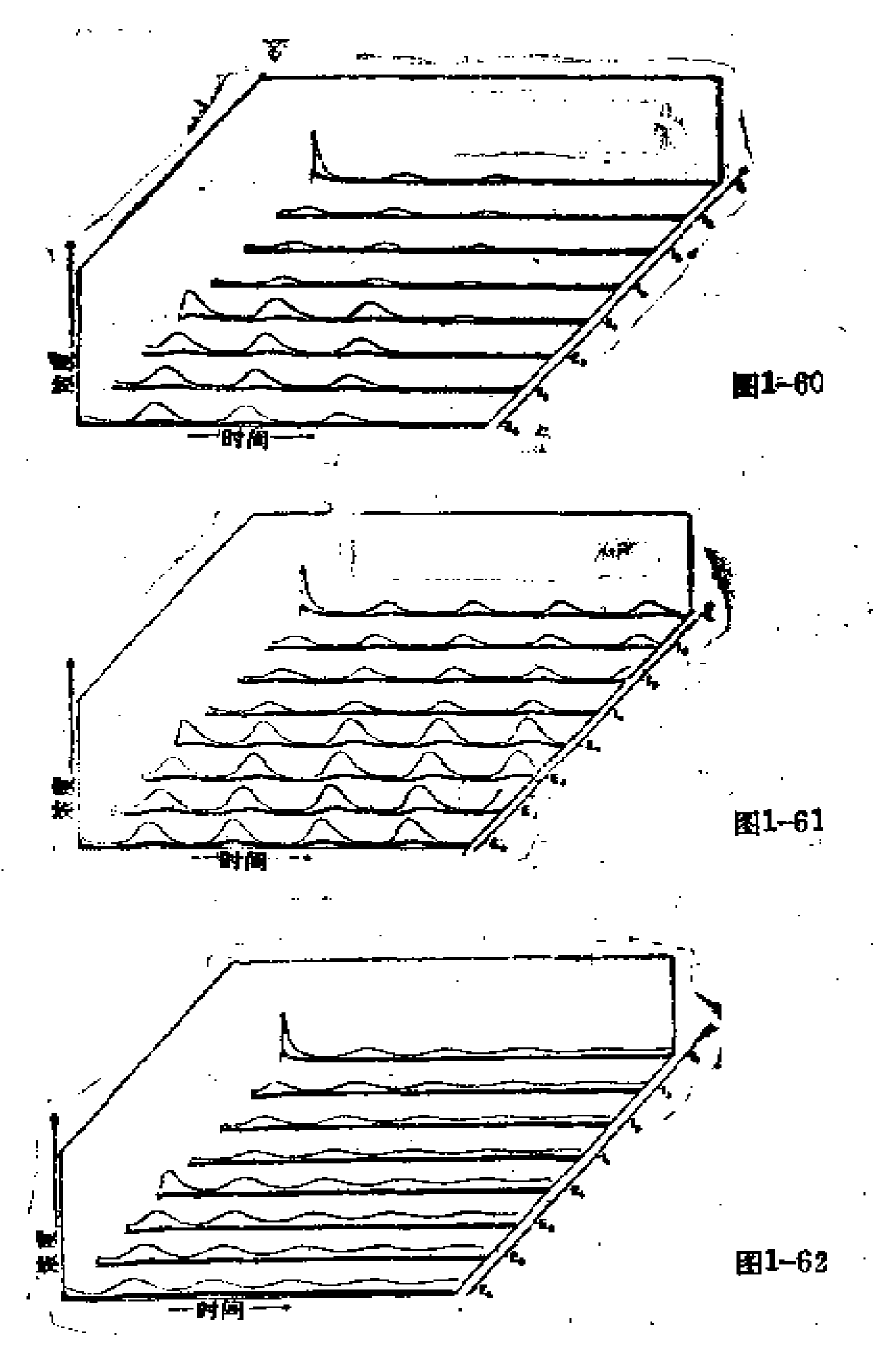


图1-60

图1-61

图1-62

图 1-60 至图 1-62　对图1-59所示的模型的偏微分方程系统进行数值积分所获的解曲线。相应于第二部分所引入的生长函数A_i的速率方程，指的是4元带翻译超循环(2.5)，仅在界面处 ($r=0$) 它有非零的催化速率项。此方程进一步考虑到吸附和解吸附(a_i 和 b_i 描述在 $r=0$ 和 $r=1$ 之间的粒子转移)、水解(在 $r\geqslant1$ 处有效)和在溶液中的扩散 ($r>1$，即在 $r=1$ 处来往于

过渡层)。每一组的3条解曲线涉及到3个空间位置$r=0$(上面)、$r=1$(中间)和$r=2$(下面)。图1-60到图1-62的差异，仅在于对于催化活性的复合物$I_i \times E_{i-1}$假定了不同的稳定常数，图1-60中最大(0.16)，图1-61中居中(0.06)，图1-62中最小(0.04)。在产生和消失之间的平衡，足以省去稀释流假设。作为所选择的值(涉及到均一的参数f_i、k_i、a_i、d_i和D，其中f_i、k_i是根据2.5的结果)的结果，在图1-60中，所有的成员I_i和E_i都消亡了，催化产生不可能与消失进行竞争，消失是由输运和分解所致。在其他两种情况中，在界面建立了稳定的超循环组织，群体数在此或是振荡的(图1-61)，或是稳定的(图1-62)。

液中超循环的行为十分相似(详细描述见第二部分)。库恩在他的论文中特别强调要考虑实际的边界条件。我们并未不同意假定"有结构的环境"，如果不能够提出实验证据以表明这样的假设至少是有用的，那么我们也不知道我们能否赞同假设非常特殊的环境。

这决不意味着把我们的模型限制在空间均一性条件(参见上述计算)。事实上，各种模型背后的逻辑推论应用于任何现实的环境，对于这些模型也就是说，存在着巨大数目的要求进行自然选择的可选择结构，由于精确度的限制，造成单个复制单元的信息量有限，或是为允许完整的系综相干地进化需要功能耦合。库恩的结论，即认为所论的此类组织是"限制在空间均一性的特殊情况"，是不中肯的。有谁在今天主张生命只能起源于多孔材料中，或起源于界面上、海水表面的复合层之中、海水里面呢？此模型表明，当且仅当一定的标准得到满足时，它才可能会以或大或小的可能性，在那些边界条件中的任一个之下起源。这些标准涉及到信息的产生和积累，应用不同的边界条件时它们没有质的差异。

对于时间均一性和不均一性，也只能说这些。在第二部分已表明了，如果把选择标准应用于定态条件，那么它们可以采

取一种特别简单的形式。因为它们与相对的反应速率而不是绝对的反应速率有关，所以不论此系统是在生长中、振荡中还是处于定态，它们在质上都是相同的。

对于许多与相分离联系的问题，缓冷的确是一种有用的程序。然而，热涨落是否对于较长的聚核苷酸的选择同样好地起作用，仍然要由实验来确定。

为了确定温度涨落是否促进了选择含较高信息量的链，人们必须仔细分析所涉及到的一切过程的相对温度系数。水解的温度系数很可能是所有系数中最大的一个。受指令的复制一般决不会在高温时得到强化。即将生成的核苷酸必须在模板上协同地结合它的互补碱基，同时利用生长链中的与顶端碱基的堆积相互作用。在模板的熔点以上这是不可能的。这些考虑适用于任何一种环境，无论它是液相、表面层，还是粗纹理或多孔的材料中的分隔间。

提高信息量的重要之处在于互补相互作用相对于非互补相互作用的相对强度。一般地，识别在低温时比在高温时更为有效。米勒和奥格尔从他们的实验数据中得出结论：

"我们不知道原始海洋中的温度怎么样，但是我们可以说，各种有机化合物和聚合物的不稳定性造成这样一种使人非相信不可的论据，即除非海洋中的温度低于25°C，否则生命不可能从海洋中产生。0°C的温度大有好处，－21°C甚至更好。在这样的低温时，原始地球上的多数水都该是处于冰的形式，只有赤道的海洋中才有海水。

"还有另一个理由使人相信，无论是在海洋还是湖泊中，生命都是在低温下进化的。所有必定导致生物组织崛起的模板－指导反应，都只发生在适当组织了的聚核苷酸结构的熔点之下。在聚核苷酸－单核苷酸螺旋的情况下，这些温度范围为0°C(甚

至更低一些)到35°C。

"生命起源的环境往往被看作是温暖的、稀淡的有机化合物汤。我们则相信，阴冷的、浓集的汤将可以对生命起源提供更好的环境。"

3.8 进化的连续性

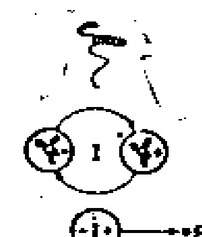
本文三部分的最后一部分的目的是要表明，超循环的确可以代表现实的物质系统，而不仅仅是我们的精神的想象产物。

进化是保守的，因此除了偶然的惊人变化以外，几乎表现为连续过程。选择的基础，实际上是优势突变体的出现所带来的不稳定性，这样的突变体打破了先前的稳定分布。不过，子代产物通常是与它们最近的祖先密切相联系的，以致变化是渐渐地出现的，前生物的进化决不违背这条规则。

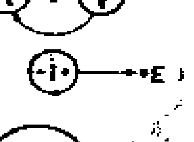
让我们把我们关于从非生命到生命转化的基本阶段的看法扼要总结一下(图1-63)。

1. 最初出现的大分子，受其结构稳定性以及它们的组分的化学丰度所支配。在早期阶段，必定存在着许多未确定的类蛋白物质以及少得多的类RNA聚合物。不过，类 RNA 聚合物从物理上继承了复制它们自身的性质，这是系统进化的必要前提。

2. 最初聚核苷酸的组成也是受化学丰度支配的。早期的核酸只不过是一类同源的大分子，其中包括 L-化合物和 D-化合物以及各种酯键，这些酯键中除了3′—5′酯键以外主要是2′—5′酯键。序列的复制性取决于复制的忠实程度。富 GC 化合物能形成最长的可复制序列。另一方面，*AU* 取代也是必要的。它们引起一定的结构柔顺性，从而有利于快速复制。可复制序



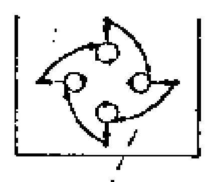
最初的聚核苷酸。



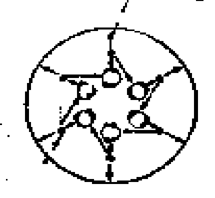
富GC拟种。



密码子分配：富甘氨酸和丙氨酸的翻译产物。

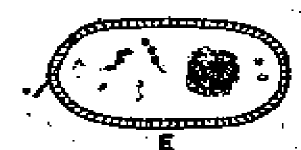


GC框架码的超循环固定，甘氨酸、丙氨酸、天冬氨酸和缬氨酸的原始复制酶的分配。



超循环组织，RNY码、复制酶、合成酶、核糖体前体的进化。密码、分隔间的进化。

完全的分隔间化的超循环，适应复制和翻译酶。代谢和控制功能、操纵子结构的进化。RNA的长度相当于今天的RNA病毒。



原细胞。
整合的基因组；DNA高级酶读出机制。进一步的达尔文进化允许分化。

图1-63　从单独一个大分子进化到整合的细胞结构的假设图解。

列形成了某种拟种分布，它显示了达尔文行为。

3．在拟种分布中的无逗点模式有资格作为信使，带有暴露的互补模式的链（很可能是信使的负链）则代表了合适的连接物。依据它们的可用性，最初氨基酸被分配给连接物。翻译产物，因为它们主要是由甘氨酸和丙氨酸残基构成，所以看起来是单调的。对于非指令的蛋白质团亦必定如此。

4．如果任一可能的翻译产物对其自己的信使复制提供了催化支持，则这种信使就可以在拟种分布中处于支配地位，并和那些与其密切关联的突变体一起以大丰度出现。此过程可以

由某些非指令性的环境蛋白质引起，这些蛋白质在其组分中反映了氨基酸的相对丰度，因此它们在性质上酷似原始指令的蛋白质。

5．按照超循环进化标准，占支配地位的信使的突变体每当提供了进一步的优势，就可以整合到复制循环中去。从而带几种密码子分配的超循环组织便能够建立起来。这种超循环组织是翻译器相干进化的一个前提。当愈来愈多的突变体整合起来，则稳步增长的精确度将允许这些序列伸长。不同的酶功能（复制酶、合成酶、核糖体因子）能够以基因复制方式从联合的前体中崛起，随后又分化了。包含几种结构基因的单元，它们一起受一种耦合因子控制。

6．复杂的超循环组织，只有它有效地利用有利的表现型变化，才能够进一步进化。为了在选择上有利于相应的基因型，空间分离（或是通过分隔间化或是通过形成复合物）成为必要，并允许在可选择的突变体组合中进行选择。形成的复合物的遗迹可在核糖体中见到。

我们并不知道，这种系统在哪一个阶段能够将其信息量完全地整合进一个巨大的基因组分子中。对此需要高度复杂的酶机构，信息存贮作用必须是一步一步地转移给DNA（这也许已发生在相当早的阶段）。

对前细胞进化的历史过程所进行的这些初步考察，已足以表明，由自复制分子单元的超循环整合所引起的发展会引向何方，以及系统最终如何敛聚为象原核细胞那样复杂的某种组织。我们想强调一下第三部分的推测性特点。自组织的早期阶段留下了痕迹，但不是留下现场情景，所以许多重要的步骤仍然是不清楚的。

我们甚至不奢望揭示出历史的真实情况。对于一个如此强

烈地依赖于偶然性的过程，在此诸如突变那样的不确定的微观事件被放大并最终决定了宏观发展的途径，完全重建历史是绝不可能的。甚至在生物学中也存在着某种“历史论的贫困”。另一方面，支配进化历史过程的原理甚至在它们更精细的细节上，也可能很好地被我们所理解。历史留在现在系统中的痕迹，可

图 1-64 对于把遗传信息的核酸的立法语言翻译成蛋白质的执行语言，遗传密码是一个普遍的线索。在这种表示中，几何空间的 3 个坐标被分配给三联体代码字中的字母位置。核酸字母表中的 4 个字母的排列方法是，丰度最大的氨基酸代码字出现在立方体的顶层。

注：*gly*＝甘氨酸　　*asp*＝天冬氨酸
glu＝谷氨酸　　*ala*＝丙氨酸
val＝缬氨酸或启动信号　　*ser*＝丝氨酸
arg＝精氨酸　　*asn*＝天冬酰胺
lys＝赖氨酸　　*thr*＝苏氨酸
ile＝异亮氨酸　　*met*＝甲硫氨酸(启动信号)
his＝组氨酸　　*gln*＝谷酰胺
pro＝脯氨酸　　*leu*＝亮氨酸
cys＝半胱氨酸　　*term*＝终止信号
trp＝色氨酸　　*tyr*＝酪氨酸
phe＝苯丙氨酸

能会提供足够的线索，以致有一天可以构造出“关于n个未知数的n个方程”。

我们想要在本部分表明的是，此类独特的反应网络即我们所谓的超循环，的确是最简单的现实的分子组织，它们保证了功能上相关的自复制单元的共存。自复制对于保持信息是必要的。因此，超循环是能够允许可复制的功能联接进化的最简单系统。它可以起源于一个自复制单元及其突变体，即起源于一种(分子)拟种。无论何时只要自然界条件允许，它的出现是不可避免的。还有："如果有谁能举出一个构成躯体构图基础而又更加美丽的三角形，我们将把他作为一位朋友而不是一个敌手而向他致意。"

(柏拉图，《蒂迈欧》[97])

与克里克、米勒和奥格尔的讨论大大地推动了这项工作，这对我们寻求分子进化中更大的连续性意味着某种“选择压力”。比布里切尔、爱泼斯坦、格特、波尔施克、西格蒙德、伍利、伍尔夫提出了特别有帮助的建议和评注。

此项工作在维也纳得到奥地利“促进科学研究基金”的赞助[项目号：3502]。

温克勒-奥斯瓦提奇设计了大部分插图，他总是一位耐心的、中肯的讨论者。

向一切提供过帮助的人们致谢。

物质的自组织和生物大分子的进化

艾 根

1 导言

1.1 "因和果"

有关生命起源的问题通常表现为"因果"问题，虽然对"因"和"果"的关系已给予了统计学的解释，但宏观过程的物理学理论通常还是要对这类问题作出回答。这主要是这个问题的本性所致，诸如活细胞这种最简单的生命形式也总是与大量复杂的宏观(即多分子)系统相联系，以致许多科学家相信，现在的物理学对生命的存在并不能提供任何明确的解释。

由于"分子生物学"这一激动人心的发现，上述问题便成了这样一个普通的问题：先有蛋白质还是先有核酸？这是古老的"鸡和蛋"疑难的现代说法。"先"的通常含义是确定某种因果关系而不是时间关系；而"蛋白质"和"核酸"这些词也可换成"功能"和"信息"。以这种形式提出的问题，当运用于诸如在活细胞中遇到的核酸和蛋白质的相互作用上，就遇到了困难；原因在于，如果没有信息存在，功能便不可能以有组织的方式出现，反之，信息又只有通过它为之编码的功能才获得意义。

这种系统可以比作一个封闭的环。显然，尽管形成环的线

条具有端点，但线条一旦封闭，起点也就无足轻重了。这一核酸和蛋白质的相互作用，相当于某种复杂的“闭环”等级组织(参见图2-1和表2-1)。要解决这种因果相互作用难题，需要一种**自组织理论**。这种自组织理论可以应用于分子系统，或更准确地说，可以应用于处于特定环境条件的特定分子系统。我们可以看到，这种分子自组织过程包括了许多没有指令功能意义的随机事件。问题在于，某些此类随机结果如何能够对其起源产生反馈，从而使它们自身成为某种放大作用的原因。在一定外部条件下，通过这种因果多重相互作用，可以建立起宏观功能组织，其中包括自复制、选择并且向高级水平进化，从而系统可以摆脱其起源时的必要条件，使环境变得有利于它自己。

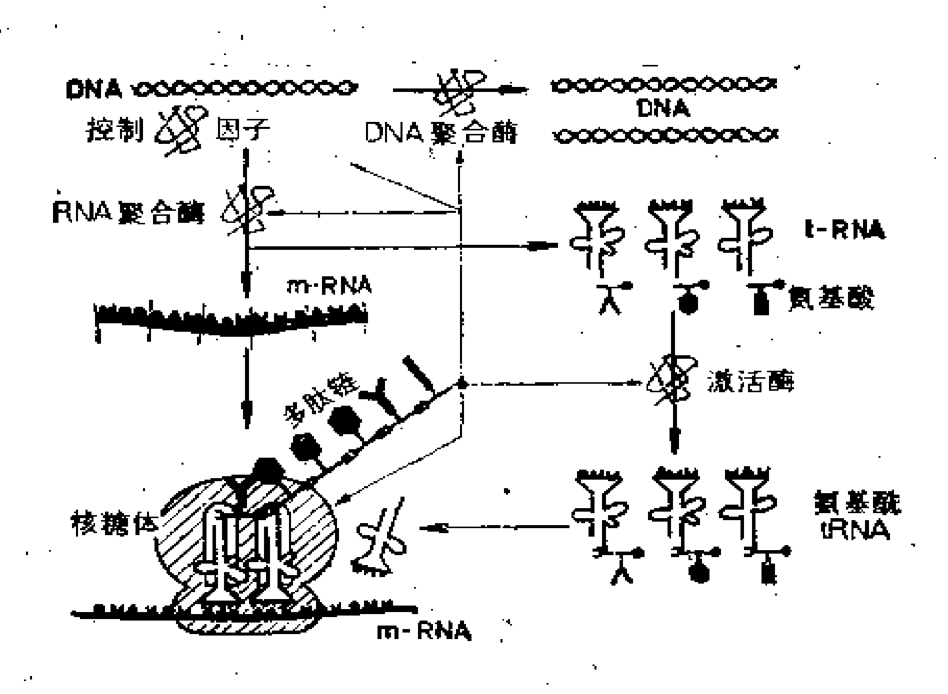


图 2-1　细胞自我复制的生物合成循环

表 2-1　核酸和蛋白质在其自复制循环中紧密连接在一起

它们的重要功能连接是:

1. DNA和 DNA 聚合酶

脱氧核糖核酸 DNA 是信息的稳定来源，它借助于 DNA 聚合酶进行复制。催化 DNA 聚合的科恩伯格酶曾被看作功能单位；但是有一些迹象表明，另有某种很可能属于膜结构的、分子量更大的蛋白质复合体，代表着细胞的实际复制机构，而科恩伯格酶还具备修复功能。

2. m-RNA 和 RNA 聚合酶

包含蛋白质综合指令的信息从DNA"转录"到一条单链的信使RNA上，取得了更"可读"的形式，这又是借助于酶实现的。(例如根据大肠杆菌)RNA聚合酶已经很好地鉴定过：分子量约为 5×10^5，其中有几个包含着特定控制因子的亚单元 $(\beta'\beta\alpha_2\omega\sigma)$。

3. tRNA 和氨基酰合成酶(激活酶)

各种氨基酸由其连接物——转移 RNA 分子进行识别，这种识别需要由氨基酰合成酶的识别部位提供"第二密码"。tRNA 是一些分子量比较小的分子(约70～80个核苷酸，参见图 2)，其组成(碱基序列)在一些情况下是已知的。氨基酸借助于其专一的激活酶(氨基酰合成酶)，在能量耦合反应中与其专一的 tRNA 连接物连接起来。这种酶是核酸和蛋白质密码之间的重要联系环节，有几个实验室在进行这方面的研究。

4. 核糖体 RNA 和蛋白质

蛋白质综合在核糖体中进行，核糖体的总分子量约为 2.7×10^6，它是由RNA及蛋白质亚单元组成的复合体，它能够容易地分为两部分。它们在超速离心机中分别沉淀为 50 与 30 个硫粒子，较小的部分具有信使的约束部位，较大的部分具有肽键形成的催化部位，两种亚单元都参与了与m-RNA 对应的氨基酸酰基t-RNA。最近已有可能将这些分裂的部分分解为单一的蛋白质及 RNA 亚单元，以形成自己的特点及模拟它们。

5. 操纵子、操纵基因、启动基因和阻遏物

转录是一种很有条理的过程，其中包括诱导和阻遏。通过蛋白质亚单元(例如 σ 因子)的调节作用，这些亚单元与 RNA 聚合酶(或其局部)协同作用，或这些亚单元是 RNA 聚合酶的一部分。乳糖操纵子的阻遏和去阻遏，这种基因控制已得到深入的研究。阻遏物是一种分子量为 150000 (4 个相同的亚单元)的蛋白质，它与 DNA 分子(约包含 10～20 个碱基对)的特定阻遏部位相互作用。通过阻遏物与(分子量较小的)诱导物的复合作用，发生去阻遏作用。我们已详细研究了这种反应机制。

1.2 自组织的前提

1.2.1 进化必定始于随机事件

无论"开端"的确切含义是什么，在"开端"处一定存在着分子的混沌，即在极其多种多样的化学物种中不存在任何功能组织。所以，与"生命起源"相联系的物质的自组织，必定始于随机事件。① 然而，这种说法并不意味着，我们今天所知道的任何有机体，哪怕是原始的有机体，能够以随机的方式拼凑起来。

若干年前，威格纳在"论自复制单元存在的概率"一文中，含蓄地作出了这种假设。他的推理实质在于：假定在量子力学意义上完全给定了某种"生命状态"以及它的滋养物和代谢产物的状态，从而能够在希尔伯特空间中用态矢来描述，于是，繁殖过程(即在有机体复制中形成的有机体与滋养物的相互作用)可以用涉及一个酉"冲突矩阵"S的变换来描述。如果假定S是一个随机矩阵，那么就能证明：代表变换的方程数大大超过未知数，即这些矢量的分矢量数；这种超过的程度，对于任何大数N和R就象N^2R与$(N+R+NR)$相比较一样。实际上，对任何现实状况，N和R是如此之大，恰如威格纳正确地断定的：如果变换方程能被未知数所满足，"那就是奇迹了"(如果许多可选择状态表示活细胞，这种说法也是成立的)。

不过，他的整个论据的基础在于如下假设：S 本质上是一

① "随机"一词，当然是指不存在功能组织，而不是指不存在物理的(即，原子、分子或者甚至是超分子的)结构。

个随机矩阵，相互作用因此是未受“指令的”，所以，与巨大的可能状态数比较，任何特定的状态几乎都是永不可能的。因此，这个结果就只能导致如下结论：任何一种我们现在称之为“有生命”的高级物质状态，都不可能是随机拼凑出来的。分子水平上出现了指令（意味着变换矩阵S必定具有某种非常特殊的形式），要求统计力学适合于特定选择和进化过程的前提。但是，这并不意味着：“现在的量子力学规律和概念必须先作修订，然后才能应用到生命问题上”。①

1.2.2 指令需要信息

我相信维纳的观点，正是他曾主张信息可以看作物理学中的一个新变量。我们现在有相当先进的“信息论”，这要归功于冯·诺伊曼、维纳、香农以及其他一些人的开创性工作。然而，我们能运用这个理论去解决自我复制难题吗？

我们现在所理解的信息论主要是一种通信理论。它所讨论的，是信息的“处理”，而不是信息的“发生”。它要求，信息“从一开始”便应以完全明确的形式存在。然后，该理论可以告诉我们，如何编码一条信息，如何利用多余度；它还能告诉我们，怎样使一条信息与信息处理机的容量相匹配，以使它通过噪声信道，从噪声中滤出，并借助编码校正装置复原出这条信息。但是，它总是需要“某种东西”——通常是人，预先定义什么是“信息”，什么是“噪声”。

这已经在信息的定义中表达出来了。在最简单的情况下，如

① 关于证明的附注。我无意在此讨论量子力学应用于“宏观过程”的一些困难。这些困难在另一些众所周知的物理过程中也已出现，虽然那些物理过程与“生命”现象没有联系。是苏黎世的普赖麦使我注意到了这一点。

果我们想从具有相等先验概率的Z_0种起始结果中选择出Z_1，那么为了把Z_0归结为Z_1种可能性，所需信息量为

$$I_1 = K\ln(Z_0/Z_1)。 \quad (1-1)$$

如果Z_1代表了一种确定的结果，即$Z_1=1$，那么所需信息量为

$$I = K\ln Z_0 \quad (1-2)$$

这是一种定义。它被用来将联合概率，即单独事件概率的乘法结合，转变成加法。因此，如果Z_0代表把λ类字符结合成ν种序列的所有可能性，那么$Z_0=\lambda^\nu$：

$$I_{\lambda\nu} = K\nu\ln\lambda \quad (1-3)$$

选择常数K使之适合于二元码，则有$K=\frac{1}{\ln 2}$。因此，如果$\lambda=2$，那么由二元字符（"比特"）的总数ν来表示，这些二元字符代表此条信息。

n个核苷酸（$\lambda=4$）的序列，其信息量为$2n$比特。把核苷酸翻译为氨基酸密码（$\lambda=20$），至少需要三联体作为编码单位。它还将允许某种多余度以及起始和终止信号。遗传密码表（表2-2）在今天已完全搞清楚了，这要归功于柯拉那、马台、尼伦格伯、奥乔亚及他们同事的工作。在第6章，我们将详细考察这个表。

表2-3颇为惊人地显示了生物大分子的巨大信息容量。在任何合理的情况下（即在大小合理的体积内），对相对短的链来说，从随机分布中再度找到任一特定序列的概率，实际上趋近于零。而这种序列的信息量还是不多的。

另一方面，一个高级生命实体的总信息量，即储存在染色体的DNA链中的总信息量，可以超过10^{10}比特，这代表了从$10^{3000\,000\,000}$种可能中的一种选择。已经有人试图把这些数字与DNA编码的结构和信息联系起来。我们随后就会明白，如果人们

表 2-2　遗传密码

		第二位置					
		U	C	A	G		
第一位置	U	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	U	第三位置
		苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	C	
		亮氨酸	丝氨酸	链终止密码	链终止密码	A	
		亮氨酸	丝氨酸	链终止密码	色氨酸	G	
	C	亮氨酸	脯氨酸	组氨酸	精氨酸	U	
		亮氨酸	脯氨酸	组氨酸	精氨酸	C	
		亮氨酸	脯氨酸	谷酰胺	精氨酸	A	
		亮氨酸	脯氨酸	谷酰胺	精氨酸	G	
	A	异亮氨酸	苏氨酸	天冬酰胺	丝氨酸	U	
		异亮氨酸	苏氨酸	天冬酰胺	丝氨酸	C	
		异亮氨酸	苏氨酸	赖氨酸	精氨酸	A	
		甲硫氨酸 [a]	苏氨酸	赖氨酸	精氨酸	G	
	G	缬氨酸	丙氨酸	天冬氨酸	甘氨酸	U	
		缬氨酸	丙氨酸	天冬氨酸	甘氨酸	C	
		缬氨酸	丙氨酸	谷氨酸	甘氨酸	A	
		缬氨酸 [a]	丙氨酸	谷氨酸	甘氨酸	G	

由 *U*、*C*、*A*、*G* 这四种碱基组成的一个三联体编码每个氨基酸，这些氨基酸都由其第一个字母标注。初看上去，每一片段中高多余度标志了三联体中第 3 个字母不甚重要。(关于第 3 个字母的取代，参见克里克的"摆动假说")人们还可见到，在垂直方向上氨基酸有一定的相似性，这标志了第 2 个字母有很大的意义。只含有*U*和 *A* 的三联体比仅含 *G* 和*C*(脯氨酸、精氨酸、丙氨酸、甘氨酸的三联体表现更多种起作用的氨基酸（苯丙氨酸、亮氨酸、酪氨酸、异亮氨酸、天冬酰胺、赖氨酸＋1个信号，所有这些事实都包含了一定的关于密码起源的信息，当然这些理论必须经受各种检验。尽管这种密码看来有普遍性，但是规定(尤其是"信号")则指的是大肠杆菌。

a．三联体 ***AUG*** 和 ***GUG*** 如果出现在顺反子的始端，指的是"起始链"或甲酰甲硫氨酸；如果它们出现在顺反子中间则分别指的是甲硫氨酸或缬氨酸。

表 2-3　基入类的 ν 个字符具有 $N\lambda\nu=\lambda^{\nu}$ 种序列

例　　子	λ	ν	$N\lambda\nu$
自然蛋白质的小亚单元（例如分子量为 1200）	20	100	10^{130}
仅从 AU 密码中形成的多肽（见表 2-2）	6	100	10^{78}
编码 33 种氨基酸的 DNA 链	4	99	10^{59}
对 AU 密码子类型的 33 种氨基酸编码的 AU 共聚体	2	99	10^{30}
含有 20 种自然氨基酸中的任何 12 种的低聚肽	20	12	4×10^{15}
含有 AU 密码类型的 20 种氨基酸的低聚肽	6	20	4×10^{15}

比较：分子量为10^4的蛋白质分子数

a）采取密堆积	
宇宙中	10^{103}
1米厚的地表层中	2×10^{40}
1立方米中	6×10^{25}
b）10^{-5}mol/L 溶液中（此浓度相当于有适当粘度的"汤"）	
所有海洋中	10^{42}
池塘中（长宽各 100 米，深 10 米）	6×10^{30}
泥潭中（1 升）	6×10^{20}

注意：多数自然氨基酸和核酸的分子量都比例子中的要大。进一步还要注意到，地球的"年龄"仅约为10^{17}秒，所以即使是快速地挨个出现蛋白质分子，人们也决不可能把所有可能的序列浏览一遍（每一个分子的寿命及组合时间肯定长于1秒）。有些数据只给了数量级。假定"宇宙"是一个直径为100亿光年的球体；密堆积指的是达到密度为1克/立方厘米。

忽视了相应实体的环境和历史，就不可能达到这一目的。进化产生的信息，是某种"有价值的"信息；关于它的功能意义，比特数是不可能告诉我们多少东西的。

上述信息定义，对于某个具有不同先验概率p_1、$p_2 \ldots p_\lambda$的字符集，需要作一些修改。多肽链中氨基酸的情况正是如此，这正如不同语言中字母的情况那样。香农认为，这里每字符的平均信息是(已选择表达式符号使之与玻尔兹曼的H函数相似)：

$$h = -K\sum_{i=1}^{\lambda} p_i \ln p_i, \quad 同时 \sum_{i=1}^{\lambda} p_i = 1 \qquad (1\text{-}4)$$

业已清楚，与相等的先验概率比较，任何约束(例如不同的多余度)都会降低每字符的平均信息量。例如，在英语中，这种约束使得每字母的平均信息量从4.76比特降到4.03比特。(诸如字母和词汇的一些优先的连续性等等其它约束，将进一步改变这个比特数。)

香农的概念与玻尔兹曼的熵的统计诠解[参见(1-2)、(1-4)式]明显相似，并且不仅仅是形式上的相似。薛定谔在其著名的小册子《生命是什么?》中，已特别强调了这一点。如果熵标征了“无知”的程度，那么任何“无知”的减少都等价于“知识”或“信息”的增加。信息和熵之间的这种互补性清楚地表明，经典信息论应用于进化问题时有局限性。信息在其有确切含义的地方，例如在约定的语言中，或者在生物学中(在进化已产生出某种固定的密码后)，这个理论就可以大显身手。然而，只要信息尚未取得其“完整意义”，或者只要对于产生出新信息还有多种选择时，这个理论就无济于事。这里我们需要一种新变量——即“价值”参数，它标征进化的水平。光是有序和无序的互补描述是不够的，而薛定谔在其小册子中对此强调得有些过分。布里渊在其优秀的专著《科学和信息论》中已明确指出，现在的信息论不适合于讨论生物学问题。

我们看到，至少对于描述无知和有知程度的熵和信息概念，“信息论”在一定程度上是对经典统计力学的补充。对于进化理

论，这种概念还是不够的。我们还需进一步阐明，知识或无知是什么。我们需要“评价”，以便表征功能序的自组织程度以及定义进化的梯度。

1.2.3 信息源于选择，或由选择获得信息

这一陈述暗示了一条基本的生物学原理，即达尔文的自然选择原理。达尔文把它看作生命的一种特殊性质：

> “我把这种有利的个体差异和变异的保存，以及那些有害变异的毁灭叫作‘自然选择’或‘适者生存’。”

他实际上对这种特征的物理学起源未作任何评断：“说科学对生命的本质或起源这个更加高级的问题还没有提出见解，这一观点是站不住脚的。”

在群体遗传学中，尤其是在费希尔，霍尔丹和赖特这些重要学派的基础工作中，已经为达尔文原理建立了数学公式。给深藏于这个原理之中的“价值概念”赋予物理客观基础遇到了困难，从而导致对这个原理的重新解释，在本文结尾处将详谈这一问题(参见8.5)。如果我们希望填平物理学和生物学之间的鸿沟，那么，我们就必须找到最终可用量子力学理论描述并为精确的分子术语解释的“选择”的意义。这样，我们就必须从已知的物质性质中导出达尔文原理。

为了把选择更紧密地与分子性质相联系并说明它的意义，让我们用蛋白质和核酸进行几场博奕。

我们利用骰子，以表示随机性：

正二十面体骰子代表蛋白质，

正四面体骰子代表核酸。

那么第一种骰子的每一面便代表了20种自然氨基酸之中的

种，而第二种骰子的每一面便代表了4种核苷酸之中的一种。此场博奕的目标在于，对ν个，比如说100个字符，循环往复地为所有字符掷骰子，并把相应字符移动到位，从而取得这些字符的某个特定序列。

若不附加任何选择规则，这种博奕将是乏味的。除了对相当小的ν以外，博奕将几乎是没有结局的。我们已经明白，一个有100个氨基酸残基的蛋白质分子，大约有10^{130}种不同的序列选择。若要取得特定序列，我们就不得不掷相应次数的骰子。这正是威格纳结论的又一个例子，它表明，若是随机地拼凑，甚至连一个有特定结构(和功能)的蛋白质分子也不可能出现。

只要相当简单地修改一下规则，我们就能在相对短的时间内完成这种博奕。对每一正确占位赋予某种"选择优势"，从而引入了"选择"。在极端的情况下，这会意味着，正确占位即是不必再为之掷骰子的位置。由于(蛋白质)在任一随机序列中有$N/20$个位置有机会被正确占据，于是现在平均只需掷$20(N-N/20)=19N$次(例如，对$N=100$，则为1900)，就足以达到此特定序列。涨落非常之大，因而使得每位博奕者获胜机会相等。这种博奕，除了领奖能激动人心以外，仍然是很单调乏味的，但是我们从中可清楚地看到试错过程的极端选择结果。要达到正确序列，所需尝试的次数只不过大约为完全指令装配的20倍。

当然，自然界进行着更为高级的博奕。"选择优势"通常并不只是一个最后的氨基酸或核苷酸残基的性质，它也并非与其他位置的发展无关。这些耦合使得博奕更能引人入胜，但也需要更多次"移子"并要求一定的策略。表4列出了一个这种"策略"博奕，在此即以核酸为例，这种博奕是相当简单的。这个例子表明：在装配$tRNA$时，为什么自然界对其二级结构倾向于采取"三叶草"结构。对核酸而言，"优势"通常不过是与是否存在互补碱

表 2-4 tRNA 博奕或如何构造三叶草

条件：每一博奕者持有一条 4 类 N 个字符的随机序列，4 类由字母 A、U、G、C 和一颗正四面体骰子来代表，骰子的每一面代表其中一类字母。

目标：轮流掷骰子，并用所获字符取代（序列中的一个位置），每一博奕者都应致力于取得尽可能多的 AU 或 GC 配对的双链结构。

规则：当某个博奕者宣称得到一种"完全"结构时，博奕终止。每一 GC 对的点数按每一 AU 对的两倍记数，得到点数最多的博奕者为胜者。约束条件是，如果不连续出现至少 2GC、或 1GC 和 2AU、或 4AU 对，则必不形成配对（协同规则）。另外，对于博奕者而言，是否形成例如发夹形、回形针形或者三叶草形等等的某种结构没有任何特定的要求，只要他对任一起始的序列服从"协同规则"。对此结构中的每一个环，他不得不至少留下 5 个未配对位置（参见图 2 的 tRNA 模型）。博奕者轮流掷骰子。每个博奕者可以随意为某个位置抛掷，但他在抛掷之前须先行宣布。另一个可能的约束条件是，要求"端—端"匹配，即两端的字符必须互相对立。

对程序的说明：人们会惊奇地发现，胜者必定以取得三叶草结构告结束，它类似于已知的 tRNA 分子的二级结构（此分子还可以有进一步的空间折迭）。

该博奕的"诡计"在于找出具有最大潜在碱基配对（包括非协同配对）数的结构。在某个具有 n 种可能配对的固定结构中，找到的概率，k 且仅是 k，可以用二项分布式来计算：

$$p_k=\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{(n-k)}\frac{n!}{k!(n-k)!}$$

对于任何固定结构，发夹形（仅具一个环）具有最大的允许配对的碱基数。不过，三叶草形有更大的灵活性，它比起发夹形来，其中结合的种类要多得多。原因在于，人们可以独立地变动其中一片叶片，从而可以获得更多种结合。这一点是肯定的。另一方面，因为有协同规则，所以不可能形成太多的叶片。对80个核苷酸而言，最佳在3 至 4 片左右（加1根梗），链更长时相应的叶片也更多。

显然，自然界早已进行了这种博奕。值得注意的是，上述协同规则与在低聚核苷酸的碱基结合中发现的稳定常数（参见第 4 节）相呼应，即 GC 配对的相互作用自由能是 AU 配对的两倍。

这种博奕的细节已由温克勒实现。在此之前，弗里斯科、阿尔伯茨和多蒂已对聚核苷酸链的最合适的二级结构作了定量估计。

基对相联系。在此核酸的转变中，互补性对合成速率和分解速

率都有巨大影响(参见第 4 章)。对于蛋白质，要把催化和控制能功（某种空间的三级或四级结构的性质）与一级序列联系起来，则要困难得多。

我们在与结构和功能有关的实验中，在这些非常罕见的酶的细节研究方面，所获知识迄今仍然很有限。

本文旨在把“选择”概念引入分子动力学中，并把它与已知的分子参数联系起来。我已经强调了一些相对平凡的博奕例子，藉以表明选择规则能够以化学性质为基础。还需要发现的是，系统将怎样利用此类结构优势？评价机制将怎样从系统的动力学行为中产生？

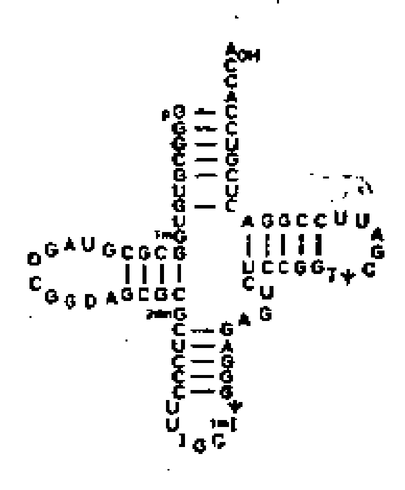


图 2-2 tRNA 的“三叶草”模型。丙氨酸专一的tRNA序列从酵母中获得。这些不平凡的碱基是：ψ=假尿(嘧啶核)苷，I=次黄(嘌呤核)苷，D=双氢尿(嘧啶核)苷，T=胸苷；1mG=1-甲基鸟苷，2dmG=N(2)-二甲基鸟苷，1mI=1-甲基次黄嘌呤苷。反密码子是5'IGC 3'(tRNA)。密码子(GCC)读作3'CCG5'(mRNA)。

1.2.4 在特定条件下特定物质出现选择

自组织的发生需要物质的什么性质？

逻辑上，我们应该把时间上并非完全割裂的进化，划分为几个时期：

1．前生物的“化学”时期

2．复制“个体”的自组织时期

3．个体物种的进化

长期以来，生物学家主要关心第3个时期。而这个时期自身又包含了许多引人注目的阶段：从性的分化和发展，通过神经细胞的发展，自我控制及通信方式的发展，终于导致了人类思维所特有的自我意识和逻辑反思。我在此无意给读者这种印象：这些阶段中的变化比起前面提到的3个时期中的变化，不那么丰富多采、特征鲜明。不过本文主要讨论第2个时期，即从“非生命”向“生命”的转变。诚如克里克在其出色的论文“分子和人”中所说：“众所周知，难于定义生命一词”；因为这种转变是非此即彼的。这样一来，如果我们要研究这一“步”，我们便应该从明确的“非生命”（例如无机物）系统入手，最终达到那些至少可以当作我们肯定地称之为“有生命”的系统，例如细菌细胞、植物和动物。

自组织以某些化学前提以及特定的环境条件为基础。它并不“正好显然”是某种物质属性。前生物时期是化学过程，因之“在原则上”可用量子力学理论加以描述。然而，我们当然应该揭示出，早期地球的条件的确有利于形成所需物质。另一个问题是，在没有任何功能序的情况下，所形成的东西比实际需要的东西多得多。化学能够解释的复杂性是惊人的，只要瞥一眼《贝尔斯登有机化学大全》就可以说明这一点。任何能够在适当条件下形成的东西都将形成，进化的第一个时期是一个分化时期，只有通过功能组织，才能使它转变成为收敛复制，并“有价值地”进化。

对前生物“化学”进化的任何详细讨论，都会超出本文的范围。在一些专著和文章中，对此有极好的综述，例如这个领域的早期开创者之一奥巴林，以及卡尔文，福克斯等人，还有庞南佩鲁马和其他人。克里克和奥格尔近期的两篇文章尤值一提，他们从实验结果出发，批评地考察了第2个时期出现的条件，本文主要关心的也正是这个时期，所有的作者都同意如下结论：生物大分子的基本建筑材料，即氨基酸、诸如ATP及其碱基同系物那样的高能核苷磷酸，以及许多其他“生物化学”化合物，能够在需要之处形成，并能够在前生物条件下即在涉及到各种能源利用的还原性大气中发生聚合。谈到这些，我并不是要制造这样一种印象，即认为涉及到这个进化时期的所有问题都解决了。许多问题依然存在，例如，3′-5′聚苷核的有序聚合过程是怎样发生的(施拉姆)及奥格尔的实验却表明了倾向其他联接，例如2′-5′，或者，活化的氨基酸怎么可以聚合成为长的多肽链而不是成为短的低聚物；这后一个问题，正如卡瑟尔斯基及其小组所表明的，显然是适当的催化条件(例如表面催化)问题。另一个问题是，生物活性物质合成所需的各种前体的丰度。所有这些问题，可能还要一代化学家继续奔忙，但是它们代表了典型的“化学”难题。

我们在此需要的，是假定象高能磷酸、激活的氨基酸等物体业已出现，并可以凝聚为显示了简单功能的大分子物体；在此，简单功能指的是，氨基酸侧链上的某些酸、碱基因的协同作用，或者增进诸如不同价态中的金属离子这类辅助因子的简单的电子给予和接受功能，或者还可象核酸那样作为简单的模板。

催化功能与各种反馈机制结合将表现为自组织的决定性前提之一这些反馈机制导致系统的某些自我促进的生长性质（参

见第2章)。按经典定义，催化剂在增加平衡速率的同时不改变该平衡。实际上，多数催化剂与底物相互作用的自由能是不可忽略的，因此要影响它们底物的平衡。不过，如果出现了任何真正的催化效应，它将以完全相同的方式作用于正、逆反应。结果是，在完全平衡或接近平衡的系统中，不可能出现“自催化生长”。这就导致了我们的第2个问题：

在什么环境条件下能够出现自组织?

薛定谔(引文同上)对此作了一个基本的回答，他写到：“生命物质避免了趋向平衡的衰退。”(孤立系统的)平衡态是熵最大的状态。如果我们要使系统离开平衡，我们就必须平稳地抵偿熵的产生，这意味着我们必须把自由能或高能物质“喂给”系统。系统的机构运用这种能量，以“推动”某些反应，使系统不致于“蜕变”为惰性的或“死寂的”平衡态。这种说法显然是正确的，薛定谔在此应受到高度评价，因为他早已如此清楚地表述了这一点。可是他也认识到，对于解释有序是“怎样”通过(其他)有序来维持以及有序是“怎样”来自无序，这种说法所起的作用是有限的。造成困难的原因在于，只把生命领域分成有序和无序是不够的。

让我们对热力学略加扩展。无论如何，为了给选择理论建立坚实的基础，我们需要这样做。梅克斯纳、普里戈金、德·格鲁特及其他人，已经发展起了不过逆过程热力学。它以表达微观可逆性的翁萨格倒易关系为基础，因而只能适用于近平衡系统。最近，格兰斯多夫和普里戈金已经把它推广到接近定态系统。我们必须加以考虑的重要的量，不是熵而是熵的时间导数 ds/dt，其中包含了两项内容，一项是与外界交换的熵流 des/dt，另一项是内部产生的熵 d_is/dt，记作σ。对于宏观系 统(运用 吉 布 斯 公式)，σ可表示为所有各项热力学流J_i与广义力X_i的乘积的和(对

非均匀系统σ应对整个系统的体积积分，为简单起见这里对此不予考虑）。近平衡系统（即流和力之间保持线性关系的系统），本质上由以下关系式表达：

$$\sigma = \sum_k J_k X_k \geqslant 0 \tag{1-5}$$

即对于任何近平衡系统，熵总是随时间增加的。平衡时熵达到极大值，此时σ等于零。

表2-5表示将它应用于化学反应系统的情况。其中，每一反应均由"反应进度"ξ_i或者它对时间的导数（即反应速度V_i，V_i=非向量流）和"亲和力"A_i（非向量力A_i/T）来标征。该例子表明，在近平衡态，σ总可由二次（正定）型的"反应进度"或者亲和力来表达。对任何近平衡系统，无论其反应级数如何，也无论不同反应之中是否存在某种耦合，在流（反应速率）和力（亲和力）之间的线性关系将保持有效。该线性微分方程系统的解是具有负实变量的指数函数。换言之，任何取其平衡值的浓度的偏离，指数都将随时间而减小：

$$(c_i - \bar{c}_i) = \sum_k a_{ik} e^{-t/\tau_k} \tag{1-6}$$

热力学平衡是以时间常数τ_k的谱为特征的"弛豫"过程。在过去20年中，已研究了许多这种相当复杂的反应系统的弛豫谱。作为翁萨格关系（意味着速率系数有对称矩阵）以及速率系数（在特征方程中导致系数仅为正的多项式）的符号的一个结果，绝不出现周期解（复指数）或者不稳定解（正变量指数）。所有的根都为负，并具有可对易的弛豫时间。

表2-5　化学热力学的定义及关系

反应i的亲和力：

$$A_i = -\sum_k \nu_{ik} \mu_k \tag{1}$$

$$=RT\left(lnK_i-ln\underset{k}{\pi}a_k^{\nu ik}\right) \tag{2}$$

计量系数:

$$\nu_{ik}\begin{cases}>0. \text{ 对反应产物}\\<0. \text{ 对反应物}\end{cases}$$

例: $4O_2+Hb\rightleftharpoons Hb(O_2)_4$,

$\nu_{O_2}=-4$, $\nu_{Hb}=-1$, $\nu_{Hb}(O_2)_4=+1$

物质 k 的化学势

$$\mu^0_k+RTlna_k \tag{3}$$

μ^0_k=标准化学势; a_k=活度(理想溶液可由浓度 C_k 代替)

平衡常数

$$K_i=\prod_k\bar{a}_k^{\nu ik} \tag{4}$$

$\bar{a}_i$ 指的是平衡态, 此时 $A_i=0$:

$$-RTlnK_i=\sum_k\nu_{ik}\mu^0_k \text{ [参见(2)式]} \tag{5}$$

反应进度 ξ_i 依据:

$$d\xi_i=dn_k/\nu_{ik} (n_k=\text{组分k的摩尔数}) \tag{6}$$

在恒温恒压下:

$$dG=\sum_i A_i d\xi_i \tag{7}$$

(G=吉布斯自由能, ξ_i 与 A_i 共轭)

反应速率:

$$V_i=d\xi_i/dt=\sum_k\varepsilon_{ik}\cdot A_k/T \tag{8}$$

(ε_{ik}=唯象系数)

翁萨格关系:

$$\varepsilon_{ik}=\varepsilon_{ki} \tag{9}$$

线性化:

$$V_i=\frac{1}{T}\sum_k\varepsilon_{ik}\sum_l(\partial A_k/\partial\xi_l)_{T,P}\delta\xi_l \tag{10}$$

矩阵(ε_{ik})和张量$(\partial A_k/\partial\xi_l)_{T,P}$ 可以写成对角型。变换后的速率方程采取如下形式:

$$V_i'=d\xi_i'/dt=\lambda_i\delta\xi_i' \tag{11}$$

式中 $\delta\xi_i'$=标准模

$\lambda_i=-\frac{1}{\tau_i}$=本征值; τ=弛豫时间

速率方程的解是

$$\delta\xi_i' = \delta\xi_{i0}' e^{-t/\tau_i} \tag{12}$$

熵产生：

$$\sigma = \frac{1}{T^2}\sum_{ik} \varepsilon_{ik} A_i A_k \geqslant 0 \tag{13}$$

表 2-6（简单的）自催化反应的例子：

$X+Y \rightleftharpoons 2X$，即借助于催化剂 X 使 Y 转变为 X①

1．远离平衡

反应速率：

$$V = k[X][Y] \tag{1}$$

在恒定的[Y]：$\delta V \sim \delta[X]$

亲和力：

$$A = RT\{\ln K - \ln([X]/[Y])\};\ \delta A \sim -\delta[X] \tag{2}$$

剩余熵产生

$$T\delta_x\sigma \to \delta V \delta A \leqslant 0 \tag{3}$$

若没有其他过程能提供某种稳定化作用，则对于恒定的供给 Y，这种系统不可能达到稳定态。

2．趋近平衡（涨落 $\delta[X]$）

反应速率

$$V = \overrightarrow{k}[X][Y] - \overleftarrow{k}[X]^2 = \{\overrightarrow{k}[Y] - \overleftarrow{k}[X]\}[X] \tag{4}$$

因为 $\overrightarrow{k}Y \approx \overleftarrow{k}X$ 且花括号项很小，因而在平衡时为零。

$$\delta V \approx -\overleftarrow{k}[X]\delta[X] \tag{5}$$

（忽略了二阶项）

亲和力：同上

剩余熵产生：

$$T\delta_x\sigma \geqslant 0 \text{（稳定平衡）} \tag{6}$$

---

① 符号 X 在此代表某种化合物，应与一般力 X_i 相区别。d 或 δ 的下标 x 指的是力的改变。

稳定性标准类似于(1–5)式，它可从爱因斯坦平衡涨落式中导出。任何在稳定平衡附近的涨落，其熵变总是小于零：

$$\delta_i s \leqslant 0 \tag{1-7}$$

对于稳定化学平衡附近的某个涨落$\delta\xi_k$，我们必须要求：

$$\delta_i s = \frac{1}{T}\Sigma A_k \delta\xi_k \leqslant 0 \tag{1-8}$$

或者，通过展开并对角线化（平衡态 $A_k = 0$）：

$$\sum_k (\partial A_k' / \partial\xi_k')(\delta\xi_k')^2 \leqslant 0 \tag{1-9}$$

因为$(\partial A_k' / \partial\xi_k') < 0$

依据一切接近真实平衡系统所共通的解的类型可知，在平衡或近平衡系统中，即使其中具有适当的物质，选择和进化也决不可能发生。

近平衡系统中，因为催化以同样的方式同时促进正、逆反应，所以自催化不可能导致生长。

处于定态的系统没有这种限制。平衡态和定态都以浓度的净变化等于零为特征，但是它们的对称关系不一样。翁萨格倒易关系不适用于定态。其结果是，在定态附近会出现振荡。

普里戈金和格兰斯多夫已经推导出一种类似于（1-5）式的关系，它对定态的邻域保持有效：

$$d_x\sigma = \sum_k J_k dX_k \leqslant 0 \tag{1-10}$$

总之，由于力X_k变化导致的熵变化（例如，在恒流下σ的变化），趋近定态时小于零，而处于定态时等于零；或者说，对于定态系统，力的某个变化所产生的熵取极小值。（注意，此关系式不适用于 $d_J\sigma$ 或 $d\sigma = d_x\sigma + d_J\sigma$） 而且，存在类似于（1-7）式的稳定性标准。这个标准认为，在稳定定态附近的任何涨落，都以某个正项反映在"产生的剩余熵"中：

$$\delta_x\sigma = \sum_k \delta J_k \delta X_k \geqslant 0 \tag{1-11}$$

这是因为在平衡态附近，熵产生的关系式（1-10）中仅含二阶项；

例如，对某个化学系统：

$$\delta_x\sigma=\frac{1}{T}\sum_k\delta V_k\delta A_k\geqslant 0 \qquad (1-12)$$

或者说：只要出现某个负涨落 $\delta_x\sigma$，定态就是不稳定的。如表2–6所示，自催化反应系统(后面将在最广泛的意义上理解“自催化”)可能出现这种不稳定性。在此速率变化归因于浓度涨落，相应的亲和力变化与其符号相反（注意，这只对于远离平衡态才正确，此时“逆”反应速率可以略而不计)。后面将看到，这种不稳定性是选择生长和进化的前提。

普里戈金及其同事已经揭示出，自催化反应行为与输运过程的结合，可以导致反应物出现某种特定的空间分布，这就是他称谓的“耗散结构”，即由能量耗散而不是守恒的分子力所形成的结构。

普里戈金认为，这些结构对于生命进化中功能序的形成具有重要意义。只要条件有利于它们的出现，它们就必定如此。所以它们也许已经对早期的形态发生某种影响，但是我不认为这些宏观结构是理解最初的生物自组织步骤的关键。这些步骤也许早就出现在没有结构的“汤”中，“汤”中肯定包含了诸如核酸和蛋白质那样的功能大分子结构。我们所需的起始阶段的自组织类型，并不是物理学(例如几何的)空间中的自组织类型。我们需要的是，在很可能是处于均相的、极其复杂多样的化合物中间的功能序。我们所需要的，是一种也许可称之为“信息空间”的空间组织。这种有序也以普里戈金和格兰斯多夫的原理为基础，对此我给予高度评价。但是，利用这个原理还将引入新的参数，从而使我们超出了现在的热力学理论范围。



唯象的选择理论

2.1 “信息”概念

具有大量各种化合物的复杂反应系统，只有在形成某种自我复制的“功能密码”条件下，才能保持有序。“功能密码”指的是两种性质，即实施性质和立法性质。实施性质需要某种机构，它能对系统中进行的所有反应进行控制，它可由相互作用和自我调节的催化剂系综来代表。这些催化剂多为均一材料制成，但是它们的功能容量实际上不受限制。无论这种实施机构的一级结构是否也对其复制提供指令，也无论这种指令是否必须从某个相异的立法源翻译过来，相关的、可复制的功能行为自组织和进化的深入，都必须从某种自我复制的分子密码水平出发。

现考虑密码载体。事实上我们知道存在着这样的密码载体，例如核酸和蛋白质，这有助于我们得出一个有用的概念。

若定义一个集合，集合元素是N_ν种信息载体$i_{\nu k}$：

$$i_{\nu k}(k=1,\ 2,\ \dots,\ N_\nu) \qquad (2\text{-}1)$$

$i_{\nu k}$以λ种碱基(例如核酸$\lambda=4$，蛋白质$\lambda=20$)的ν个字符的序列为特征。为简单起见，我们往往只涉及一类均一的信息载体，即

长度相同的、有特定的ν个字符数的一类序列。

在经典信息论中，某特定序列$i_{\nu l}$的"信息量"由比特数来表示：

$$I_\nu = \nu ln\lambda / ln_2 [\text{比特}] \quad (2-2)$$

那么，N_ν代表了字符数为ν的这类序列的"结构容量"，即N_ν是可能形成λ种碱基的长度为ν的序列的总数：

$$N_\nu = \lambda^\nu = 2^{I\nu} \quad (2-3)$$

如果我们考虑到任一长度都以ν个字符为上限，我们得到的不同序列为：

$$\sum_\nu N_\nu = \frac{\lambda^\nu - 1}{1 - \lambda^{-1}} \quad (2-4)$$

如果我们考虑相互独立的竞争系统（包括从1到ν的任何长度），其中每条序列代表一条（具有不同"选择价值"的）信息，那么这个序列数就有着重要的意义。

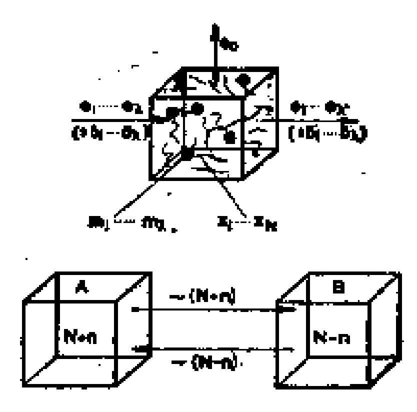


图 2-3 **"信息匣子"** 匣子具有半透壁，高能和低能的单体字符都能经它出入。在匣子内部，代表"信息"的聚合序列是可集聚和可解聚的。模板指令的集聚和解聚两者都可以是酶促过程。通过控制流量（$\phi_1, \ldots, \phi_\lambda; \phi_1', \ldots, \phi_\lambda'$）或控制浓度（$m_1, \ldots, m_\lambda$；由稀释流$\phi_0$调节$x_1, \ldots, x_N$），定态可以得到保持。为进行比较，由两个匣子$A$和$B$来代表平衡态，此时系统的分布$N_A = N_B$，还包括涨落$\pm n$（参见3.2中讨论的埃伦菲斯特的缸子模型）。

信息载体系统可以置于某个有限体积的“匣子”中，高能和低能的单体字符能通过匣子的可透壁进出(参见图2-3)。在匣子中，每种信息载体可以(在每单位体积中)有$x_{\nu i}$个拷贝，这样一类序列的总数为：

$$n_\nu=\sum_{k=1}^{N\nu}x_{\nu k} \tag{2-5}$$

对于N_ν种可能的序列$i_{\nu k}$，其浓度$\nu_{\nu k}$将为零。匣子中未组织的高能单体字符1，…，λ，其浓度为$m_1,\dots,m_\lambda$，它们流入匣子的流量为ϕ_1，…，ϕ_λ。

已组织的及未组织的字符总数为

$$M_0=\sum_{\lambda}^{\lambda}m_k+\sum_{\nu}\nu n_\nu \tag{2-6}$$

匣子中的“组织化程度”D_0，即已组织的字符与字符总数的比

$$D_0=\frac{\sum\nu n_\nu}{M_0} \tag{2-7}$$

几乎在一切实际情况下，它将都表明，在体积V中的一类序列的群体数，比起N_ν的信息容量来，都将小得多：

$$n_\nu\cdot V\ll N_\nu \tag{2-8}$$

(例如，对$\nu=100$，$\lambda=20$，如同10^{20}或更小一些的数与10^{130}相比较，参见表2-3)。一个重要的结果是：没有自我指令时，它是一种随机分布，对某一给定序列的期望值实际上等于零。而且，对那些偶然形成的序列，再度偶然找到这个序列的一个相同拷贝的概率实际上也是零。对进化的最优方案而言，这种“未饱和的”信息容量的性质，具有极其重要的意义。

我们进一步的任务是要确定信息载体的某些动力学性质，并建立起选择理论。这个理论应该包括导出一个用分子术语表达的“选择优势”参数。

为处理已耦合的系统，我们可设想出由群体变量x_v以及其中的功能关系的集合所定义的“信息空间”。真正的线性系统不可能“选择”，对准线性系统，这可以导致一种类似于处理线性驰豫现象的“标准模式”。

2.2 唯象方程

给定某类信息载体 i_{vi}，其中每种载体在每单位体积中有 x_{vi} 个拷贝：

$$0 \leqslant x_{vi} < n_v \quad (n_v \cdot V \ll N_v) \tag{2-9}$$

在达尔文意义上的选择，必定涉及该系统的动力学性质，这些性质由系综的速率和信息载体的分解来表示。当然，我们必须从大分子物种的发生和转变的速率方程入手，因为大分子是进化行为的代表。

对远离平衡的系统，我们不能期望流和力之间存在着简单的线性关系。这样，在不可逆过程热力学理论的线性范围内(例如，在速率与亲和力之间的简单关系)，从唯象方程的形式入手是不利的。由于已揭示了选择需要某种相当专一的反应行为，所以运用能更清楚地说明所讨论的反应类型和级数的化学速率理论的术语，也许更为有利。

我们提出 3 个假设，它们将表现为选择的必要前提：

1．系统必须是开放的，并远离内平衡。为了阻止系统向平衡衰退，我们必须平稳地输入自由能，例如输入高能单位(诸如 *ATP* 及其碱基系列物，或活化的氨基酸)。另一方面，分解将产生出低能产物。所以，信息载体的生成和分解两种反应，都由正的亲和力所推动；这两个过程不象近平衡态的生成和分解，这

里不存在任何"微观可逆"关系。

2．生成速率必须大于分解速率，而且至少在$x_{\nu i}$中有相同的级数。由于在$x_{\nu i}$中的分解通常（至少）是一级的①，该生成必定具备某种"自催化"本性。如果在$x_{\nu i}$中生成速率比分解速率的级数小，该系统就不具有抵制较低效率竞争者的选择所需的内在生长性质。在这种情况下，所有的$i_{\nu i}$将只能生长到某个恒定的水平，此时它们的生成速率与分解速率"相匹配"。这种系统将携带前面突变的所有无用信息，从而终将阻止任何进一步进化。

3．由于条件$n_{\nu}\cdot V \ll N_{\nu}$，任何个体信息载体的非指令生成都完全可以忽略不计。

在第4至第6章将表明，自催化行为包含了许多不同种类的反应，其中只有某些类型才谈得上具有进化行为。阐明大分子信息载体的"自我指令"行为与"一般催化"行为的区别，是极其重要的。一个自催化过程，其间可能通过涉及到若干个中间物的某种反应循环（参见第5章），使得反应的产物对其自身的生成具有反馈。某个给定的$i_{\nu i}$，其生成速率也许不与它的浓度$x_{\nu i}$成比例，而是与某一信息载体的浓度$x_{\nu j \neq i}$成比例，而比值$x_{\nu j}/x_{\nu i}$最终达到某个恒定值。在此，自催化是某特定系综的特殊性质。需要"自我指令"的信息载体具有一般的"模板"性质。这就意味着，任一字符序列都将指导自己的复制。如果在复制中出现某个误差，这种误差便会重新复制出来。这是核酸的典型表现，而"一般催化"行为在蛋白质中也能发现。

---

① 在酶促分解过程中，如果酶已被饱和，那么反应速率可以变得与底物浓度无关；不过，情况并不总是这样。底物在其生长期间总是要通过某"非零的级数范围"，而且在饱和条件下的竞争仍然涉及到个体物种的群体变量。（例见第7章）

出现某个信息载体的唯象速率方程，现在可写成一般形式($\dot{x}\equiv dx/dt$)①：

$$\dot{x}=(F_i-R_i)x_i+\sum_{l\neq i}\varphi_{il}x_l \qquad (2\text{-}10)$$

其中的第1项和第2项分别指的是信息载体i_{v_i}的自我指令的生成和消失(例如，由分解、稀释等等)。第3项包括所有进一步的(非自发的)产生项，它们起源于其他极其相似于i_{v_i}序列的不精确复制，F_i和R_i是一般的速率常数，其中可以包含若干个项。选用特殊的(2-10)式的形式，是为了表达出"内在的"自催化要求，它不必意味着是一级反应行为。速率参数F_i肯定是单体字符浓度($m_1, \ldots, m_\lambda$)的函数，而且F_i和R_i都可能进一步依赖于x_i或其他物种的群体变量x_k。

我们可以用纯唯象术语进一步详细说明F_i和R_i(这里可能包含各个反应项的和)②：

$$F_i=k_0A_iL_i,\ R_i=k_0D_i+\varphi_{ii} \qquad (2\text{-}11)$$

k_0代表了一般的一级反应速率常数，具有量纲秒$^{-1}$。对所有竞争的生成过程，它定义了合适的恒定参考值(或阈值)。于是其余的各个参数A_i、L_i和D_i都不含时间量纲。

乘积项A_iL_i描述了各个不同生成速率参数F_i的大小和形式。A_i被称之为"放大因子"；k_0A_i实际上是一个速率常数，它描述由模板i_i指导的合成有多快。因此，k_0A_i涉及由模板i_i指令出现的(单位时间的)所有生成过程，而不论它们是导致i_i的准确拷贝还是突变体。L_i是"品质因子"，反映了这些过程中有L_i部分导致i_i的精确拷贝。由i_i指导的所有

---

① 如果我们不打算明确指明相应的类，则后面一律省去下标。

② 选择这些特殊的符号，是为了强调速率和品质参数的一般形式，这些参数可以是浓度的复合函数，并包含几个项。

(译者注：原文在以后公式中的F、A、L、R、D都是用花体字。)

"复制"过程的分数 $(1-L_i)$ 则反映了突变体 i_j 的生成，i_j仍然部分地类似于主宰拷贝 i_i，但其中包含了以某种概率分布出现的各个误差。这些突变体的特征通常是速率参数 $F_j \leqslant F_i$，但有时也会出现 $F_j > F_i$，即出现有利的拷贝。

A_i 和 L_i 的定义纯粹是唯象的。人们可以记下由一定模板指令的(每单位时间的)复制过程次数，并由序列分析或其他分析来决定正确生成的物种 i_i 的数目；这就使得两个参数都有了确定的物理意义。如果我们考虑总产生速率，则可表达为：

$$\sum_{k=1}^{N} F_k x_k + \sum_{k=1}^{N} \sum_{l} \varphi_{kl} x_l = k_0 \sum_{k=1}^{N} A_k x_k \tag{2-12}$$

右项不再包含有品质参数 L_i，因为产生的总体包括了正确复制部分L_i以及误差(或突变体)的产生部分 $(1-L_i)$。实际上，(2-10)式中的最后一项解释了：任何误差的出现必定在密切有关的拷贝的产生项中表现出来。对于特定的被选择物种，这个项可以很小或者甚至可以略而不计，但是对一个被选择的主宰拷贝，其"误差子代"的可复制生成则可能具有重要意义。在没有任何自发的、非指令的合成情况下，由守恒要求我们得到关于误差产生的总和：

$$k_0 \sum_{k=1}^{N} A_k (1-L_k) x_k = \sum_{k=1}^{N} \sum_{l} \varphi_{kl} x_l \tag{2-13}$$

乘积 $A_i L_i$ 中也包含计量函数 $f_i(m_1, \ldots, m_\lambda)$，$f_i$ 描述了生成速率对单体(高能)字符浓度的依赖性。它们的精确形式取决于聚合过程的特殊机制(参见第4章)，特别是其中的限速步骤。如果浓度 $m_1, \ldots, m_\lambda$ 是缓冲的(多数进化实验都要选用的条件，亦见第4章)，f_i 便可作为一恒定因子被包括在$A_i L_i$中。否则，我们必须说明：

$$A_i L_i = f_i(m_1, \ldots, m_\lambda) A'_i L'_i \tag{2-14}$$

(并且可能还相应于不同误差拷贝的平均值)。$A'_i L'_i$ 不再包含 $m_1, \ldots, m_\lambda$的浓度项，但是仍然可以包含 x_i或者任何 x_k的浓度项。[$f_i(m_1, \ldots, m_\lambda)$ 可以标准化，例如对于起始或终止条件便可这样做。]

由R_i 的关系式 (2-11)，我们区分独个分解项 D_i (仍然与一般速率常数k_0相联系)与稀释项 φ_{0i}，φ_{0i} 在速率方程中也与x_i 相联系("成比例

的"稀释)。如果这种比例是明确的，我们可以省去下标 i，并使 φ_0 与总稀释流 $\phi_0$①相联系：

$$\varphi_0=\phi_0/\sum_{k=1}^{N}x_k \tag{2-15}$$

总消失速率可以表示为：

$$\sum_{k=1}^{N}R_kx_k=k_0\sum_{k=1}^{N}D_kx_k+\phi_0 \tag{2-16}$$

正如前述，参数 D_i 可以是浓度 x_i 或 x_k 的函数；但是在此我们不可能区别任何"品质"参数，因为任何一个物种的分解都会导致无用的产物("废物")——如果我们不引入更高级的修复机制。

由效应而不是错误复制所引起的突变体，能够在形式上包括在这3个参数 A_i、L_i 和 D_i 之中。

(2-10)式现采取如下形式：

$$\dot{x}=R_0[A_il_i-D_i]x_i+\sum_{l\neq i}\varphi_{il}x_l-\varphi_{0i}x_i \tag{2-17}$$

这个唯象速率方程一般地描述了由如下性质分类的任何反应系统：

a) 代谢，由两个总速率项 $R_0\sum D_Rx_R$ 和 $R_0\sum A_Rx_R$ 表示，这两项描述了高能材料向低能材料的转变。

b) 自复制，如速率方程的形式所表达。假定生成和分解速率都正比于 x_i，在忽略掉 F_i 和 D_i 之间进一步的浓度关系条件下，对于 $0<x_i<n$，有 $F_i>D_i$。

c) 突变性，由品质因子 $L_i<1$ 表示。它表明，(2-17)式的第一部分对于选择行为取定值，A_i、L_i 和 D_i 是决定性的唯象参数。甚至对复杂的"生命"实体，选择也是由这些参数来决定。由于"内部"耦合，这些参数可以包含复杂的浓度项，并依赖于

① 根据系统的"浓度"，φ_0 的符号可以为负。然而，φ_0 项通常是用来补偿生长的。

许多环境变量。

但是，(2–17)式这种形式不能描述选择过程。它定义了某种分离。由于阈值性质：

$$A_iL_i \gtrless D_i \tag{2–18}$$

因而在阈值以上($A_iL_i > D_i$)的物种将生长，而在阈值之下($A_iL_i < D_i$)的物种将消亡。

如果我们暂不顾 (2-17) 式的第二部分，(2-17) 式的解可以一般地写作：

$$x_i(t) = x_i^0 exp\left\{k_0\int_0^t (A_iL_i - D_i)\, dt\right\} \tag{2-19}$$

对于恒定的 A_i、L_i 和 D_i，式中出现具有正变量或负变量的实指数。如果这些反应中有某个反应接近平衡态，方程的解就会取得具有负变量的指数形式，正如1.2.4中表明的。它来源于可逆(一般为非线性的)速率项的展开式，作为变量，它偏离了平衡态的"反应范围"(参见表2-5)。

非线性系统可以表现出更明显的"分离"行为。例如，如下类型的微分方程：

$$\dot{x} = (A_iL_i - D_i)x_i，式中\ (A_iL_i - D_i) = a_i + b_ix_i \tag{2-20}$$

如果 b_i 和 $(a_i + b_ix_i^0)$ 大于零(其中 a_i 值甚至可以为负)，在无限大的 t 处产生了奇点。对于 $x_i^0 = x_i(t=0)$，我们有：

$$x_i(t) = x_i^0 \frac{a_ie^{a_it}}{a_i + b_ix_i^0(1 - e^{a_it})} \tag{2-21}$$

当

$$t = a_i^{-1} ln\left(1 + \frac{a_i}{b_ix_i^0}\right), \tag{2-22}$$

$x_i(t)$ 便趋于无限。如果a_i为负，其绝对值则必须小于$b_ix_i^0$，否则这个解衰退。如 $a_i \ll b_ix_i^0$，这个解将还原为简单的双曲线：

$$x_i(t) = \frac{x_i^0}{1 - b_ix_i^0t} \tag{2-23}$$

且奇点在 $t = 1/(b_ix_i^0)$。

这些解及类似的解，提供了相当明显的选择行为(参见第 6 章)，并对最初的进化产生了极其重要的意义。

我已倾向于把前面所描述的行为称作"分离的"行为，而不是称作"选择的"行为，这是因为它仅仅导致系统分离成为由阈值性质($A_iL_i \gtrless D_i$)所区别的两部分。如果我们想通过某种极值原理来解释"选择"，那么我们不仅需要单种组分的生长性质，而且需要某些"外部的"选择约束，以强迫系统进入真正的生存竞争。

2.3 选择约束

我们可以设想出许多外部约束条件及内部耦合作用（参见沃尔特拉对"斗争"问题的处理），它们能使反应系统更富有竞争性。不过，有两个直接的程序将为选择理论提供一般的基础。这两个程序都可以与普里戈金和格兰斯多夫对定态邻域中反应过程的处理联系起来。

如果我们迫使系统保持某个定态，按照热力学理论(参见第 1 章)，则可以认为，这种系统或是处于恒反应力，或是处于恒反应流。

至于在2·1中引入的信息匣子，类似地，我们可以保持字符的总组织（因而就有某种总亲和力，参见表2–5）恒定，或保持字符的总流恒定（由单体字符的进出流量以及匣子的总反应流来确定）。

特别地，第一个条件指的是：已组织的和未组织的字符总数以及匣子中的"组织度"必须保持恒定。要从物理上做到这一

点，可以缓冲高能单体的浓度 m_1，…，m_λ，并以保持信息载体总数不变的方式(通过ϕ_0)来控制总流量。

对恒总组织，约束条件为

$$m_1,\ \dots,\ m_\lambda = \text{常量} \tag{2-24}$$

这还意味着，$f_i(m_1,\ \dots,\ m_\lambda)=$常量，并可以包括在 A_iL_i 中：

$$\sum_{k=1}^{N} x_k = \text{常量} = n \tag{2-25}$$

(如果考虑不同类的 ν，则 $\sum_\nu \nu n_\nu=$常量)

必须调节稀释流量ϕ_0，以抵偿总的超额产量：

$$\phi_0 = k_0 \sum_{k=1}^{N} (A_k - D_k) x_k \tag{2-26}$$

在另外的选择条件下允许匣子中含量发生变化，但是，单体高能材料的流入量(ϕ_1，…，ϕ_λ)是不变的，反应流即信息载体(还包括低能的分解产物流出量)的总转变速率也是不变的：

$$\phi_1,\ \dots,\ \phi_\lambda = \text{常量} \tag{2-27}$$

或者，如果单体有相同的来源时，

$$\phi_M = \sum_{k=1}^{\lambda} \phi_k = \text{常量}$$

以及

$$k_0 \sum_{k=1}^{N} A_k x_k = k_0 \sum_{k=1}^{N} D_k x_k = \phi_M = \text{常量}。 \tag{2-28}$$

无论在实验上还是在理论上，对后一种情况进行处理都将更为困难。自然界的某些过程，可以非常接近这些条件。例如，在某个周期内，可以恒定速率供给能量(例如，恒定流入太阳能)，以调节高能原料的水平，从而产生出某个恒定产生速率。(速率参数的增加由单体字符的浓度水平减小来抵偿。)类似地，信息载体可以生长到某个水平，此时它们的分解被调整到与其生成速率相等。不过，精确保持这种条件需要高级的控制。对进化

实验而言，保持恒总组织条件要容易一些（参见第7章中西格尔曼用噬菌体Qβ 进行的系列转移实验）。人们可以建造自动控制并保持该特定条件的“进化机器”，还可以想象包括上述两种条件的各种组合的其他约束条件。

注意：为理解进化原理和对实验资料进行定量评价，对这些确定条件详加说明是重要的，但是，自然界的任何真实进化过程不并一定在这些特定条件下才能发生，这正如从来没有热机是在卡诺循环精确规定的热力学平衡条件下进行工作的。

现在让我们回到唯象方程，并在考虑如下两种不同选择约束条件的情况下重写该方程。

1）恒总组织

并相应把下式称为：

$$E_i = A_i - D_i \qquad \text{（超额）生产力} \qquad (2\text{-}29)$$

$$\bar{E} = \frac{\sum_{k=1}^{N} E_k x_k}{\sum_{k=1}^{N} x_k} \qquad \text{“平均生产力”} \qquad (2\text{-}30)$$

$$W_i^0 = A_i L_i - D_i \qquad \text{“选择价值”} \qquad (2\text{-}31)$$

所有这些量都涉及到恒总组织的约束。

按照(2-15)和(2-26)式代换 φ_0，我们可把(2-27)式改写成如下形式：

$$\dot{x} = k_0(W_i^0 - \bar{E})x_i + \sum_{l \neq i} \varphi_{il} x_l \qquad (2\text{-}32)$$

相应于某个“误差修复”或某个“信息逆流”的项 $\sum_{l \neq i} \varphi_{il} x_l$，对于任一被选择的主宰拷贝 i_m，通常可以略而不计。可是，总有一条错误拷贝的“彗星尾巴”跟随着这个“主宰拷贝”，错误拷贝的平

稳出现主要归因于生成项 $\varphi_{im}x_m$（下标 m 指的是被选择的复本）。由于每一群体变量 x_i 出现在平均生产力 $\bar{E}$ 中，所以(2-32)式本质上是非线性的，即使对恒值的 A_i、L_i 和 D_i 亦如此。所以，所有方程都通过这个项发生耦合，这就提供了某种反映系统自组织可变动的并可自我调整的阈值。信息载体只有当其"选择价值"W_i^0 在阈值 $\bar{E}$ 以上才可以生长。作为它们生长的结果，阈值 $\bar{E}$ 将平稳地变大，直至达到最佳的 $\bar{E}$ 值，此时它相应于所有物种的最大选择价值：

$$\bar{E} \to W_m^0 \tag{2-33}$$

或者更一般地（例如，在振荡系统中），对任何振荡周期 Δt：

$$\int_t^{t+\Delta t} \bar{E}dt \to \int_t^{t+\Delta t} W_m^0 dt \quad 或 \quad \langle \bar{E} \rangle \to \frac{1}{\Delta t}\int_t^{t+\Delta t} W_m^0 dt \tag{2-34}$$

按照反应系统的类型（即在各种组分之中的耦合），如果参数 A_m、L_m 和 D_m 是恒定的，W_m^0 将属于某一个（或某一简并类）信息载体；或者，它可以涉及到某个催化循环，从而是循环中所有成员的浓度函数，或者，它甚至可以包括由更高级的浓度项所表达的反应循环的整个等级组织。对于非线性系统，任一物种最终的 W_m^0 值，要取决于浓度的初始条件，这正如(2-24)式的例子所表明的。重要的是要注意到，W_m^0 中的下标 m 指的是：相对于在最后阶段出现的所有竞争者，物种 m 具有"最大的"选择价值。而且，关系式(2-33)或(2-34)在"误差修复"项可忽略不计时仅仅是一种近似。否则，人们可用下式代替 W_m^0：

$$W_m^0 + \frac{1}{k_0}\sum_{l \neq m} \varphi_{ml} \frac{\bar{x}_l}{\bar{x}_m},$$

式中，第二项代表了来自突变体的平均信息"逆流"（这也可由特定的修复功能项来修订）。

对于(2-33)式的更高级近似，小修复项的贡献体现在：

$$\overline{E} \longrightarrow W_m^0 + \frac{1}{k_0^2} \sum_{l \neq m} \frac{\varphi_{ml}\varphi_{lm}}{W_m^0 - W_l^0} \tag{2-35a}$$

只要第 2 项相对于任何$(W_m^0 - W_{l \neq m}^0)$充分小，这个近似就保持有效。如果L_i非常接近 1，则总是可以保证这种充分小的。

这样一来，有几分类似于热力学平衡，我们用某个极值原理标征了选择。不过，在热力学平衡中，我们得到的是“真正的”熵的极大值或者自由能的极小值。我们在此则仅仅处理“最优值”，即在某些约束条件下相对于可选择性组成的极大值。我们也可以把这种由(2-33)或(2-34)式作为标准所标征的状态，称作“选择平衡态”，不过我们应该认识到，我们在此讨论的是亚稳平衡态。当涨落的误差分布中(参见 $L<1$及$n \cdot V \ll M$)不出现“选择优势”时，它稳定了群体中“适者”复制的信息；“选择优势”是由$W_{m+1}^0 > W_m^0$所标征的。只要出现了这种新拷贝(或其系综)i_{m+1}(参见第 3 章中的随机处理)，先前的平衡态就会崩溃，新的(亚稳)平衡态就将出现。它是以不同的$\overline{E}$为特征的，它的总体变化相当于一个优化过程。如果W_m^0与任何x_i无关，那么$\overline{E}$的变化(在恒定的环境条件下)相应于单调增加：

$$W_m^0 < W_{m+1}^0 < \cdots < W_{最优}^0 \tag{2-35b}$$

終态是一个“最优态”，即在不等式形式给出的约束条件下的一个极大值。系统只能取得所有N种可能集合状态中的某些状态；它被限定在某个必须“攀登”的途径上。如果出现了由与浓度相关的参数W_i^0所表达的内部耦合，那么整个最优化过程就更为复杂。群体中的最大选择价值并无“绝对的”意义，因为它现在指的是浓度x_i的分布。在这种分布中，x_i分布的任何变化都代表了“环境的变化”。最优化在此涉及到“不同的”过程；它可以正好伴随着(可能是所有的)选择价值的某种一般的减小(例如，作为由某个被选择的信息载体引起的“污染”的结果)。它亦可以包

括利用更多的信息量(以便对付变化了的环境)。一般说来，最优的进化程序不一定要求选择价值的某种简单的单调变化。不论终态是什么，$\bar{E}$ 在此都是指在最后阶段出现的所有竞争者中的最大W^0值。协同系统的物种，它们都终将达到相同的W^0_m值，这个等式关系可以用于计算被选择物种的“平衡”分布，这类似于“质量作用定律”。

重要的是注意到，“选择”(出现在固定环境条件下的一定的群集状态的集合中)与“进化”(作为进一步改变群体和环境的优化过程)的区别是一种抽象。如果把整个进化过程看作一场博奕，那么，对于评价按某些规则进行的博奕状态，这种抽象把选择机制当作了某种实施工具，从而代替了博奕者。仅仅对 $n \cdot V \ll N$ 且 L 接近于1的系统，这种抽象才近似正确。它还要求，实验条件要保证出现选择比之进化变化所需的时间要短。

b) 已组织和未组织的字符的恒总流量

比起涉及到第一类约束的系统，唯象方程系统的形式更复杂。因此，让我们考虑一个简化的例子，这个例子仍具有一切本质特征！假定所有的信息载体(包括错误拷贝)都具有接近一致的总组成成分，它们的生成速率可以用相同的(平均)计量函数 $f(m_1, \ldots, m_\lambda)$ 来描述。而且，假定高能单体(以恒定比例)的流量 ϕ_M 是均一的、恒定的。引入(3-27)和(3-28)式的约束条件，则导致

$$f(m_1, \ldots, m_\lambda) = \frac{\phi_M}{k_0 \sum_k^N A_k' x_k} \tag{2-36}$$

并且，类似于(2-32)式可写出速率方程(忽略了“逆流”项)：

$$\dot{x} = \frac{k_0 D_i}{\bar{P}+1}[W_i^0 - \bar{P}]x_i \tag{2-37}$$

式中 $\bar{P}$ 是指某种“生产力”的平均值，不过，其定义：

$$\bar{P}=\frac{\bar{A}^{1}}{\bar{D}}-1 \tag{2-38}$$

式中：

$$\bar{A}^{1}=\frac{\sum_{k=1}^{N}\bar{A}_{k}^{1}x_{k}}{\sum_{k=1}^{N}x_{k}}$$，$\bar{D}$ 与之类似。

生产力和选择价值为：

$$P_{i}=\frac{A_{i}^{1}}{D_{i}}-1,\quad W_{i}^{F}=\frac{A_{i}^{1}L_{i}}{D_{i}}-1 \tag{2-39}$$

这里涉及到恒流约束条件。

如果 $f(m_1, \ldots, m_i)$不表示某个普遍有效的计量函数，那么人们仍可除开各个计量项，并运用合适的平均数而形式地获得(2-36)式。

对于应用第1章结尾处已讨论过的普里戈金和格兰斯多夫原理，这种恒流情况是特别有趣的。他们的原理涉及到在恒定流量时力的变化。由于改变了阈值，该系统又要选择在现存群体中的最大选择价值。这里可以表明，每一次导致"选择价值"的某种进一步增加的突变，相应于代表现存定态不稳定性的熵产生的负涨落。在恒流条件下的进化相应于这种不稳定性的结果，其中，依据一种有限的(正的)选择优势($W_{m+1}^{F}-W_{m}^{F}$)，占支配地位的物种i_m将消亡，从而有利于新物种i_{m+1}。

2.4 选择平衡

我们已把一定群体的最大生产力状态，称作"选择平衡态"。这种"平衡态"中有误差分布的涨落；相对于出现具选择优势的物种，它是亚稳的。然而，如同在化学热力学中那样，我们能

够从(2-33)式、(2-34)式，或者类似于恒流约束等条件下，分别导出“平衡常数”。

a）恒总组织

我们把平均生产力$\bar{E}$写成：

$$\bar{E}=\frac{E_m x_m+\sum\limits_{k\neq m} E_k x_k}{\sum x_k}=\bar{E}_{k\neq m}+\frac{x_m}{n}(E_m-\bar{E}_{k\neq m}) \qquad (2-40)$$

其中残余生产力的平均值定义为：

$$\bar{E}=\frac{\sum\limits_{k\neq m}\bar{E}_k x_k}{\sum\limits_{k\neq m} x_k} \qquad (2-41)$$

这里运用了

$$\sum_{k=1}^{N} x_k=\text{常量}=n,\quad \text{以及}\ \sum_{k\neq m} x_k=n-x_m \qquad (2-42)$$

在一级近似下（$L\lesssim 1$，忽略“逆流”项，参见2.6.c），对于被选择物种的“平衡分数”，由平衡条件(2-33)式可得到：

$$\frac{\bar{x}_m}{n}=\frac{W_m^0-\bar{E}_{k\neq m}}{E_m-\bar{E}_{k\neq m}} \qquad (2-43)$$

选择标准可写作：

$$W_m^0>\bar{E}_{k\neq m}$$

这就从物理上定义了达尔文的“适者”一词。

我们注意到，虽然生存率$\bar{x}_m/n$不正比于L_m，但是平衡误差分数$(1-\bar{x}_m/n)$正比于$(1-L_m)$，即：

$$1-\frac{\bar{x}_m}{n}=\frac{A_m}{E_m-\bar{E}_{k\neq m}}(1-L_m) \qquad (2-44)$$

假若L_m等于1，则W_m^0等于E_m且$\bar{x}_m$趋近n。这是极端的选择过程，它对继续进化毫无用处。这种系统已获得的“价值”，被限定从所有n种物种中的相对有限的（随机的）一些物种中进行选择。我

们还发现，除非我们说明了"对于什么"，否则"价值"一词就毫无用处。对于特定环境下的选择，其价值不同于更一般的最优进化的价值。对于后者，我们需要的是一个等式因子L_m，它充分大因而足以保证生存：

$$L_m > L_{min} = \frac{\overline{A}_{k\neq m} + D_m - \overline{D}_{k\neq m}}{A_m} \qquad (2-45)$$

但是，另一方面，L_m又要尽可能小，以提供尽可能多的突变体，从而能够出现继续"进化"。该系统必须满足这个非常重要的选择条件(2-45)，以保存已获得的"信息"。因此，一个小的L_m($>L_{min}$)，其效应是双重的：一方面，它产生着更加多种多样的突变体，以便系统能从中选择，从而允许更大的极限最优值W_m；另一方面，它加大了进化速率。

在耦合的反应系统中，选择平衡涉及到整个信息载体系综的生存。我们在此必须求解代数方程系统。"选择价值"以及"生产力"的表达式可以变得相当复杂，其中包括了该系统的所有A、L和D参数。在第5章和第6章将讨论一些例子。对于线性系统，我们可以变换变量，用其行为类似于单个自我指令物种的浓度"标准摸"来代表整个系综。而且，对于某些非线性系统，可以详细算出其平衡关系。显然，对那些非线性系统，"平衡"一词与通常的化学平衡有很多共同之处，因为它与几种组分有关，有时甚至与许多组分有关。不过，它们是有差别的，在真正的平衡中各种物质互相转变；而在选择平衡中，各种物质从一种集合状况分解成为另一种集合状况，且两者之间不能互相转变，但是只要保持环境条件不变，它们的分解就以一定的固定比例进行。

b) 恒信息流

类似于(2-43)式，我们运用2.3.b引入的定义，可以计算被选择物种的平衡率，结果得到(又是作为对 $L \lesssim 1$ 的一种近似)：

$$\frac{\bar{x}_m}{\phi_m/k_0 D_m} = \frac{\bar{x}_m}{\bar{n}} = \frac{W_m^F - \bar{P}_{k+m}}{P_m - \bar{P}_{k+m}} \tag{2-46}$$

这里 $\bar{x}_m$ 并未标准化为 $\sum_{k=1}^{N} x_k$，与恒总组织的情况不同，它不是不变量，而象一个常量(在给定的 D_m)：

$$\frac{\phi_M}{k_0 D_M} = \frac{\sum_{k=1}^{N} D_k x_k}{D_m} = \bar{n} \tag{2-47}$$

流量 ϕ_M 在此是守恒量，$\phi_M/k_0 D_m$ 类似于某种总浓度。而且，与恒组织的选择平衡相类似，由 $W_m^F > \bar{P}_{k+m}$ 给出了选择标准。

对于选择价值的3个参数 A、L 和 D，只要确定其中一个，就可以引入选择优势。相应地，可以区别出3种极端情况(参见表2-7)，其中包括抑制、去抑制、特定的促进、特定保护阻滞的消化，等等。

表 2-7　选择标准(关于 A_m、L_m 和 D_m 的极限情况)

1. $D_m = D_{k+m}$，即 $\phi_M/R_m = \sum_k x_k = n$

   恒力和恒流

   $\frac{\bar{x}_m}{n} = \frac{A_m L_m - \bar{A}_{k+m}}{A_m - \bar{A}_{k+m}}$；若 $A_m L_m > \bar{A}_{k+m}$，则选择。

2. $A_m = A_{k+m}$

   a) 恒力

   $\frac{\bar{x}_m}{n} = \frac{(\bar{D}_{k+m} - D_m) - (1 - L_m)\bar{A}_{k+m}}{\bar{D}_{k+m} - D_m}$；

   若 $\bar{D}_{k+m} > D_m + \bar{A}_{k+m}(1 - L_m)$，则选择。

   b) 恒流

   $\frac{\bar{x}_m}{n} = \frac{L_m \bar{D}_{k+m} - D_m}{\bar{D}_{k+m} - D_m}$；若 $L_m \bar{D}_{k+m} > D_m$，则选择。

3. $A_m L_m = \bar{A}_{k+m}$

   a) 恒力

$$\frac{\bar{x}_m}{n}=\frac{\bar{D}_{k\neq m}-D_m}{(A_m-\bar{A}_{k\neq m})+(\bar{D}_{k\neq m}-D_m)};\ 若\ \bar{D}_{k\neq m}>D_m,则选择。$$

b) 恒流

$$\frac{\bar{x}_m}{n}=L_m\frac{\bar{D}_{k\neq m}-D_m}{\bar{D}_{k\neq m}-L_mD_m};\ 若\ \bar{D}_{k\neq m}>D_m,则选择。$$

2.5 品质因子和误差分布

显然，“选择价值”作为一种动力学性质依赖于诸如 A 和D那样的速率参数。而品质因子L的作用则不甚明显，这个因子描述复制的精确性。在简单的模型中，L可以明确地与诸如单字符识别精确度 q 这样的分子参数联系起来。对于单字符的匹配(互补)配对和错误匹配(非互补)配对的相互作用，可以根据自由能(或激活参数)来测量 q_i。通常，这种识别中有协同作用，对其中的最近(可能还有次最近)相邻配对需作详细说明。这将大大扩大可能的组合数。第 4 章将讨论一个关于核酸碱基 A、U、G 和C($\lambda=4$)的无酶识别的例子。

如果只要区别两种稳定性，即一种是所有的(简并)互补配对以及另一种是所有的(简并)非互补配对，那么，唯象参数 L 和分子(单字符)识别参数 q 之间的关系是相当明确的。表2–8列出了这种例子的相互关系。它们可以推广到包含若干类非简并的字符(例如核酸)。品质因子L 的意义，可用图 4 所示的误差分布清楚地揭示出来。误差的期望值越小，相应的L 曲线就越尖锐，即“品质”因子越高。重要的进化标准(2–45)式定义了一个最大信息量(ν_{max})，它将最小品质因子 L_{min} 与速率参数的“扩展”相联系，对于一给定的识别参数q，ν_{max} 能复制地得到保持[参见表2–8中的(10)式]：

$$\nu_{max}=\frac{|\ln L_{min}|}{1-q}$$

作为一个结果，基本的识别机制(即 q) 必随着进化期间某物种的信息量I_ν的增加而得以改进。在分子进化的较高水平上，要求相当高级的控制机制，以改进单字符识别的精确度（例如酶的密码核实），并允许 A 和 D 参数充分传播(参见第 4 到第 6 章)。

表 2-8 "双态"模型的品质因子 L

精确复制一个单字符的概率：q

$(\lambda-1)$重简并误差其中一个出现的概率：

$(1-q)$

在 ν 个字符的序列中的误差期望值：

$$\varepsilon=\nu(1-q) \tag{1}$$

简并性：1 无缺陷的排列

$\nu(\lambda-1)$ 单缺陷的排列 (2)

$\binom{\nu}{k}(\lambda-1)^k$ k 缺陷的排列

和：$\sum_{k=0}^{\nu}\binom{\nu}{k}(\lambda-1)^k$ 条可能序列(参见表 2-4) (3)

出现无误差拷贝的概率：

$$L=Q_{\nu 0}=q^\nu\longrightarrow e^{-\varepsilon} \tag{4}$$

仅出现 k 缺陷的序列的概率分布

$$Q_{\nu k}=q^{(\nu-k)}(1-q)^k\binom{\nu}{k}\text{(二项分布)} \tag{5}$$

对 $k(1-q)\ll 1$，由泊松分布近似：

$$Q_{\nu k}=\frac{\varepsilon^k\cdot e^{-\varepsilon}}{k!} \tag{6}$$

错误拷贝的和

$$\sum_{k=1}^{\nu}q^{(\nu-k)}(1-q)^k\binom{\nu}{k}=1-q^\nu=1-Q_{\nu 0} \tag{7}$$

在确定的位置出现包含 k 误差的一个特定突变体的概率：

$$P_{\nu k}=\frac{Q_{\nu k}}{\binom{\nu}{k}(\lambda-1)^{k}}=\frac{q^{(\nu-k)}(1-q)^{k}}{(\lambda-1)^{k}}\longrightarrow\left[\frac{q^{-1}-1}{\lambda-1}\right]^{k}\cdot e^{-\epsilon} \tag{8}$$

一定突变体(例如具有某种选择优势)的产生速率与 P_{ν} 成比例。

概率分布(5)、(6)和(8)如同图2-8所示。

对于稳定选择，$Q_{\nu 0}$ 的最小值已由(2-45)式定义。对某个给定的识别因子 q，稳定选择的标准可以表示为：

$$\nu=\frac{|\ln L_{\min}|}{|\ln q|} \tag{9}$$

或者对于 $(1-q)\ll 1$：

$$\nu=\frac{|\ln L_{\min}|}{1-q} \tag{10}$$

定义了可以正确复制并保存的信息量的极限值 $I_{\max}$。

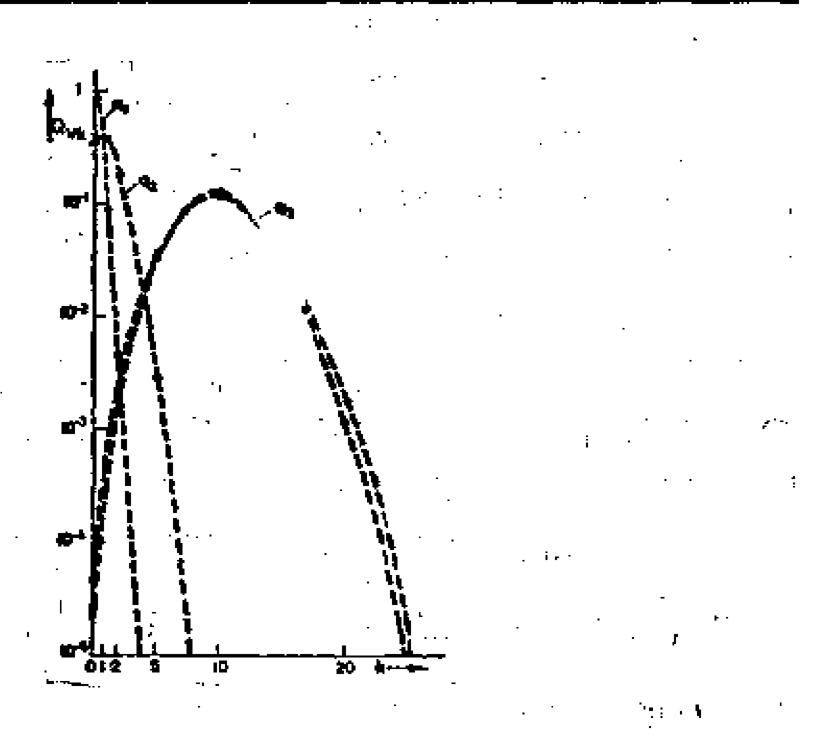


图 2-4　出现误差复本的概率分布：$Q_{\nu k}$ 依据表2-8中的(5)式

参数是：

$\nu=100$	$q_1=0.999$	$\varepsilon_1=0.1$
$\lambda=4$	$q_2=0.990$	$\varepsilon_2=1$
	$q_3=0.900$	$\varepsilon_3=10$

q_3 的两条曲线表示(5)式的二项式分布与(6)式的泊松分布之间的比较。

q_1 以及 q_2 的这两种曲线，对小的 k 值（在作图的精度内）是重合的。然而，对 $k \longrightarrow \nu$，则出现较大的偏离。曲线的极大值出现在 $k=\varepsilon$ 处。如果以线性尺度作图，则分布是极其尖锐的。相应于(8)式的 $P_{\nu k}$ 没有图示出来，因为随着 k 的增加它下降得极快。例如对于 q_2:

k	$P_{\nu k}$
0	3.7×10^{-1}
1	1.2×10^{-5}
2	4.2×10^{-6}

这种讨论的目的在于，揭示出品质因子L_i不仅对于给定群体中的选择，而且对于继续进化的速率以及最优终态都是重要的。系统力图在由相当大的速率参数标征的群体中取得较高的L值，而这似乎不利于继续进化的速率。另一方面，较高L值允许生成的物种具有较大的信息量，这最后又将有利于继续进化。由此可见，最优化过程常常可以包括着互相矛盾的要求。例如，由酶进行的专一底物识别，对于酶—底物复合物来说需要高的稳定常数，而过高的稳定常数则限制了转变速率。这个由价值概念的推广所引起的最优进化的问题，将在第8章讨论。在此我们已经明白，"价值"总是需要对被评价的性质进行详细说明，评价越详细，则进化的水平越高。

这些考虑的基本结果是，必须复制保存最大信息量 $I_{max}\left(=\nu_m\frac{\ln\lambda}{\ln 2}\right)$，必须适应于基本字符识别的精确度。这个由参数 A 和 D 的变化所定义的量，只以对数项出现，因此只对小的 A 和 D 的变化(即 $A_i\approx\bar{A}_{k\neq i}$ 以及 $D_i\approx\bar{D}_{i\neq k}$)才有限制作用。

2.6 选择动力学

对于两种约束条件，这些唯象方程都是非线性微分方程系统。当然，显式解取决于由特定反应机制所确定的方程的特殊形式。第4到第6章将讨论几个包含蛋白质和核酸的自组织系统的机制。这里我们仅讨论恒量参数W_i的一些原型解，以便于找出选择过程的特征。让我们考虑3种增长复杂性的情况：

1．恒总组织，$\bar{E}_{k\neq i}$可变； $L_i=1$，即$W_i^0=E_i$

只要突变拷贝的信息流交换（由$\sum\limits_{l\neq i}\varphi_{il}x_l$表示）可以完全略而不计（$L_i=1$），就可以给出（2-3）式的微分方程系统的精确解。方程系统具有简单的形式：

$$\dot{x}=k_0(W_i^0-\bar{E})x_i \tag{2-48}$$

方程的解$\{x_i^0=x_i(t=0)\}$

$$x_i(t)=\frac{x_i^0 n\,exp(k_0W_i^0t)}{\sum\limits_{k=1}^{N}\frac{x_k^0E_k}{W_k^0}exp(k_0W_k^0t)} \tag{2-49}$$

式中任一W_k^0均可用E_k替换。

从(2-48)式的积分所获隐式出发，也可得到这个解：

$$x_i(t)=\frac{exp(k_0W_i^0t)}{exp\left\{k_0\int_0^t\bar{E}(\tau)\,d\tau\right\}} \tag{2-50}$$

对任何可被插入$\bar{E}$的比(x_k/x_i)，积分项可以去掉，这时$\bar{E}$写作：

$$\bar{E}=\frac{x_i}{n}\sum_{n}^{N}\frac{x_k}{x_i}E_k \tag{2-51}$$

速率方程(2-48)变为：

$$\dot{x}_i = k_0 W_i^0 x_i - \frac{\sum_{k=1}^{N} x_k^0 E_k exp[k_0(W_K^0 - W_i^0)t]}{n x_i^0} x_i^2 \tag{2-52}$$

它代表了伯努利微分方程的一种特殊形式：

$$\dot{x} + g(t) + f(t)x^2 = 0 \tag{2-53}$$

它具有众所周知的解：

$$\frac{1}{x(t)} = e(t)\int \frac{f(t)}{e(t)}dt, \quad e(t) = exp(\int g(t)dt) \tag{2-54}$$

插入：

$$g(t) = 常量 = -k_0 W_i^0 \tag{2-55}$$

$$f(t) = \frac{k_0}{n x_i^0} \sum_{k=1}^{N} x_k^0 E_k exp[k_0(W_k^0 - W_i^0)t] \tag{2-56}$$

于是得出解(2-49)。

这个解清楚地描述了一种选择程序。在 $t=0$，每一 x_i 均由其起始值 x_i^0 给出。对 $t \to \infty$，指数和可由最大项表示，它属于具有最高"选择价值"W_m^0 的物种。正是这种物种将得到选择。它趋近于定常值：

$$\bar{x}_m = \frac{W_m^0}{E_m} \cdot n \tag{2-57}$$

只要 $Q_m = 1$(即 $W_m^0 = E_m$)，它就等于 n。所有其他物种都将按照下式衰退：

$$x_i(t) = n \cdot \frac{x_i^0}{x_m^0} \cdot \frac{W_m^0}{E_m} \cdot exp[(W_i^0 - W_m^0)k_0 t] \tag{2-58}$$

在 x_m 生长到占有"支配"水平前，一些 $x_i(t)$ 在衰退前也许已在起始的增长中通过了极大值。图 2-5 显示了 4 种竞争物种的例

子。

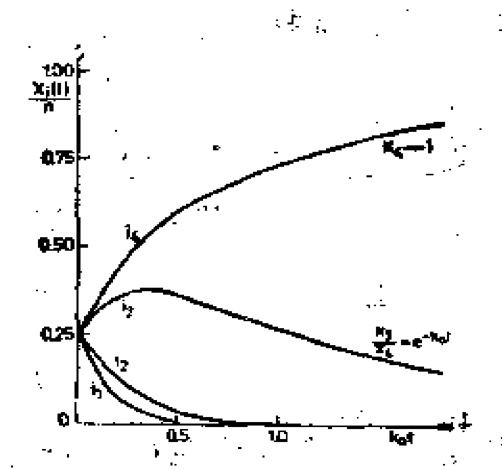


图 2-5 在4种竞争物种中的选择 依据(2-40)式，且 $W_1^0=1$，$W_2^0=4$，$W_3^0=9$，$W_4^0=10$。

如果 Q_m 并不正好等于1，而是接近1，那么这种处理仍然可对主宰拷贝提供有用的解。当被选择的“主宰拷贝”已生长到占有支配的水平，它将主要与从不正确复制中形成的突变体竞争。前述的处理并不能对这些突变体的行为提供解释；按(2-13)式，对它们必须考虑到附加速率项(尤其是$\varphi_{im}x_m$)。结果是，它们的浓度将不会衰退到零。对被选择的主宰拷贝，只要 $\bar{E}_{k\neq m}\ll W_m^0$，(2-49)和(2-57)式就提供了一个良好的近似。在此我们不打算详细说明任何错误拷贝，只是要认识到存在着一定量的误差。(如果 $W_m^0=E_m$ 即$Q_m=1$，该系统便不能继续进化。)在对主宰拷贝和突变体之中的相互作用详加考察之前，我们可以讨论一个特例，它对 $L_i\lesssim1$ 可直接给出一个解。

2．恒总组织或恒流，$\bar{E}_{k\neq m}$或 $\bar{P}_{k\neq m}$ 为恒量

这种情况或者描述两物种 m和k 之间的竞争，或者描述与具有恒定 $\bar{E}_{k\neq m}$ 或 $P_{k\neq m}$ 的 $k\neq m$ 的物种的一个简并类的竞争。

我们还看到，对于在 $E_{k\neq m}$ 或 $P_{k\neq m}$ 中并不简并，但在平均值 $\bar{E}_{k\neq m}$ 和 $\bar{P}_{k\neq m}$附近表现出相当恒定分布的物种，这种情况代表了一种良好的近似，而被选择的物种具有$W_m^0>\bar{E}_{m\neq k}$或$W_m^F>\bar{P}_{k\neq m}$的明确的复制率。

在这种恒总组织的情况下，被选择物种的解为：

$$x_m(t)=x_m^0\frac{exp[(W_m^0-\bar{E}_{k\neq m})k_0t]}{1+\frac{x_m^0}{\bar{x}_m}\{exp[(W_m^0-\bar{E}_{k\neq m})k_0t]-1\}} \qquad (2\text{-}59)$$

式中

$$\bar{x}_m=n\frac{W_m^0-\bar{E}_{k\neq m}}{E_m-\bar{E}_{k\neq m}}\text{［参见(2-43)式］}$$

这是在 2.4 中引入的、在 $x\to\infty$ 时取得的 x_m 的“平衡”值。而且，所有具有$W_{k\neq m}^0<W_m^0$的“独立”物种终将衰退为零，而所有突变体（依据假定的 $\bar{E}_{k\neq m}=$常量，它们是简并的）的和则将达到由(2-44)式所表示的定常水平。

对于L_m还有一种限制：它不能太小，以便来自突变体的物种 I_m 的任何再现（通过在相继复制中的误差反转）都可略而不计。L_m 的容限在此取决于信息量，即取决于所涉及的字符数 ν。据(2-43)式，这种限制对“平衡”值仍然有效。另外，在一个群体的最后选择时期，实质上只有在突变体（和主宰拷贝）之中的竞争，解式(2-59)在此是一种良好的近似。平均值$\bar{E}_{k\neq m}$在此的确近似看作一个常量，所以这个解也得出 x_m 的正确的“平衡”值（对选择“平衡”的邻域也保持有效）。另一方面，第1种情况，即(2-49)式，除非$\bar{E}_{k\neq m}\ll W_m^0$，否则不会导致正确的平衡值；但是在选择的起始时期，突变体数比无关的拷贝数少一些，此时(2-49)式是一个良好的近似。$L=1$ 时类似于(2-59)式的解，在群体遗传学的文献中已得到反复讨论。一个众所周

知的例子是罗斯的疟原虫扩散方程。洛特卡对此类方程的解的稳定性给出了一般的函数理论标准。

在恒流量的约束下，至少假定所有物种有一致的计量函数 $f(m_1,\dots,m_i)$，则这一假定与(2-36)和(2-46)式相联系便可以明确地求解这个"两物种"竞争问题。在速率方程(2-37)式中，$\bar{P}_{k\neq m}$、W_m^r 和 $\bar{A}=n\bar{D}=\phi_M$ 等项是常量(采取简单的一级生成和分解速率)，而 $n\bar{A}^1$ 项以及 $(1+\bar{P})=\bar{A}^1/\bar{D}$ 是 x_m 的线性函数。个体的速率方程有如下形式：

$$\dot{x}=\frac{A+Bx_m}{C+Dx_m}x_m \tag{2-60}$$

其中

$$A=\phi_M D_m(W_m^r-\bar{P}_{k\neq m});\quad C=\frac{\Phi_M}{k_0}(1+\bar{P}_{k\neq m})$$

$$B=-k_0 D_m^2(P_m-\bar{P}_{k\neq m});\quad D=-B/k_0 D_m$$

均可以进行积分，并得出解：

$$\frac{[x_m(t)]\alpha_m}{\bar{x}_m-x_m(t)}=\frac{[x_m^0]\alpha_m}{\bar{x}_m-x_m^0}\cdot exp\{k_0 D_m(1-\alpha_m)t\}$$

或者可选用

$$\frac{[x_m(t)]^{\alpha_m}}{[x_m(t)]^{\alpha_m}+[\bar{x}_m-x_m(t)]}=\frac{[x_m^0]\alpha_m exp\{k_0 D_m(1-\alpha_m)t\}}{(\bar{x}_m-x_m^0)+(x_m^0)\alpha_m\cdot exp\{k_0 D_m(1-\alpha_m)t\}} \tag{2-61}$$

式中

$$\alpha_m=\frac{1+\bar{P}_{k\neq m}}{1+W_m^r}\quad 或\quad (1-\alpha_m)=\frac{W_m^r-\bar{P}_{k\neq m}}{1+W_m^r} \tag{2-62}$$

这个解的性质十分相似于(2-59)式。对 $t\to 0$，"被选择的"物种按下式生长：

$$x_m(t)=x_m^0 exp\left\{\frac{k_0 D_m}{1+\bar{P}_{k\neq m}}(W_m^r-\bar{P}_{k\neq m})t\right\} \tag{2-63}$$

[对 $x_m(t)\ll\bar{x}_m$ 进行了近似]；而对于 $t\to\infty$，$x_m(t)$ 趋近于"平衡值"：

$$\bar{x}_m = \bar{n} \cdot \frac{W_m^r - \bar{P}_{k\neq m}}{P_m - \bar{P}_{k\neq m}}$$ [参见(2-46)式]

3、恒总组织，$\bar{E}_{k\neq m}$可变，近似考虑误差产量：$L<1$(但不是$\ll 1$)

在被选择的物种生长期间，从主宰拷贝 i_m可复制地出现的突变体将在总群体中不断增多。依据缺陷位置的数目（与主宰拷贝比较而言），即依据 $\sum_k i_{1k}$、$\sum_k i_{2k}$ 等等，可对这些突变体进行分类。对 ν 个字符的序列，每一类中的拷贝数由二项系数给出；对 λ 类字符，在给定位置的每一个缺陷是$(\lambda-1)$重简并的；这样一来，具有 l 个缺陷的一类，有$\binom{\nu}{l}(\lambda-1)^l$ 种不同的拷贝。不同的各个拷贝 i_{lk} 的产生频率，将随 l 的增加而减少，例如，对于假定具有一致的 q 因子的最简单模型，单字符识别(参见表2-8)与下式成比例：

$$\frac{q^{(\nu-l)}(1-q)^l}{(\lambda-1)^l}$$

相应地，每一类缺陷对主宰拷贝的恢复作出贡献，并将随缺陷数增加而减少。

为了定量地了解一下突变体对选择的这种影响，让我们考虑一种简单近似：单字符缺陷的影响。仅仅当误差$\nu(1-q)$的期望值适当小于1，即$0\ll L<1$，这种近似才有效。

然后，我们必须区分3类速率方程（$W\equiv W^*$ 总是涉及到恒组织）。

a）对一个主宰拷贝 i_m：

$$\dot{x}_m = k_0(W_m - \bar{E})x_m + k_0 \sum_{k=1}^{\nu(\lambda-1)} \frac{E_{1k} - W_{1k}}{\beta_{1k}} x_{1k} \qquad (2\text{-}64a)$$

b）对具有一个单字符缺陷的$\nu(\lambda-1)$种突变体 i_{1l}：

$$\dot{x}_{1i}=k_0(W_{1i}-\bar{E})x_{1i}+k_0\frac{E_m-W_m}{\beta_{mi}}-x_m \quad (2\text{-}64b)$$

c) 对 $(n-x_m-\sum_k x_{1k})$ 种独立的竞争者 i_{ji}：

$$\dot{x}_{ji}=k_0(W_{ji}-\bar{E})x_{ji} \quad (2\text{-}64c)$$

第3类独立竞争者，也可以看作具$j\geqslant 2$个缺陷的主宰拷贝的突变体。对充分长的序列，这个可能的竞争者数是如此之大，以至偶然找到任何指定拷贝的概率实际上为零。可以假定，它们作为主宰拷贝的突变体是很少的，可以略而不计。这包括那些其主宰拷贝只在两个位置有缺陷的突变体。这种近似要求(只要独立竞争者对该解有实质贡献)：

$$\sum_k x_{2k}\ll\sum_{j>2}\sum_k x_{jk} \quad (2\text{-}65)$$

在(2-64a)和(2-64b)式中的因子β，一般地使特定突变体的产生与总缺陷的产生相联系，这由下式表示：

$$A_i(1-L_{i0})=(E_i-W_i)$$

对 $L_{i0}=q^{\nu}$ 的一致的 q，以及对一致的误差产生率(例如 β_{mi})归结为

$$\beta_{mi}=\nu(\lambda-1) \quad (2\text{-}66)$$

因为从主宰拷贝产生的所有误差的分数 $1/\{\nu(\lambda-1)\}$，相应于任何特定的单缺陷拷贝的产生。

为了获得自洽的近似，我们必须忽略掉除了主宰拷贝中单缺陷产生以外的任何突变体生成，以及忽略掉从单缺陷拷贝再生出的主宰拷贝。这就需要假定，$W_{ji}\approx E_{ji}$，$(E_{1k}-W_{1k})=A_{1k}(1-q)$ 以及 $\beta_{1k}=1$。后一个假定，即对任何单缺陷拷贝假定 q 取代 L，初看起来似乎有些不现实，这是因为它仅仅允许一种错误，即导致该主宰拷贝的恢复。不过，忽略掉单缺陷拷贝复制中的其他误差(它们会导致双缺陷的拷贝)是与这种近似融洽的，并仅仅影响小的校正项。对任何单拷贝 i_{1k}，E_{1k} 与 W_{1k} 之差肯定很小，可忽略不计；然而在(2-64a)式中出现的所有这些项的和，在这种近似中则必须加以考虑。

对主宰拷贝，我们获得如下与时间相关的浓度$[x_m^0=x_m(t=0)]$：

$$\frac{x_m(t)}{n}=\frac{W'_m\bar{x}^0_m}{x^0_m(E_m+\sum_{k=1}^{\nu(\lambda-1)}A_kE_{1k})+\sum_{k=1}^{\nu(\lambda-1)}B_k exp\{k_0(W_{1k}-W'_m)t\}+\sum_{j,k}C_{jk}exp\{k_0(W_{jk}-W'_m)t\}} \tag{2-67}$$

式中：

$$A_k=\frac{E_m-W_m}{\beta_{mk}(W_{1k}-W'_m)}$$

$$B_k=\frac{W'_m}{W_{1k}}E_{1k}\left(x^0_{1k}+x^0_m\frac{E_m-W_m}{\beta_{mk}(W_{1k}-W'_m)}\right) \tag{2-68}$$

$$C_{jk}=\frac{W^0_m}{W_{jk}}\cdot E_{jk}x^0_{jk}$$

W'_m 能够表示为：

$$W'_m\approx W_m+\frac{E_m-W_m}{W_m-\bar{E}_{1k}}\left\langle\frac{E_{1k}-W_{1k}}{\beta_{1k}}\right\rangle \tag{2-69}$$

式中，符号〈　〉代表对所有$\nu(\lambda-1)$个单字符复本的“平衡平均值”。可以发现，对$L_m\to 1$，W'_m还原为W_m，对$t\to\infty$，该解趋近于平衡比率[参见(2-43)式]：

$$\frac{\bar{x}_m}{n}=\frac{W'_m-\bar{E}_{1k}}{E_m-\bar{E}_{1k}} \tag{2-70}$$

在眼下的近似限度以内，需要主宰拷贝由充分大的选择优势$W'_m>W_{1k}$来区别，式中的任何残余项都极小，可省略不计。

W'_m与W_m之差可以从一致的字符识别模型看出来，该模型导出：

$$W'_m-W_m\approx\frac{\nu(1-q)^2}{\lambda-1}\cdot\frac{A_m\bar{A}_{1k}}{W_m-\bar{E}_{1k}} \tag{2-71}$$

式中$(1-q)\ll 1/\nu$。(注：此模型中的平均值$\bar{A}_{1k}$和$\bar{E}_{1k}$代替了先前的平均值$\bar{A}_{k\neq m}$和$\bar{E}_{k\neq m}$。)

对任一单缺陷拷贝i_{1l}的解，以及对独立竞争者i_{jl}的解，都由下式与主宰拷贝的解相联系：

$$\frac{x_{1l}(t)}{x_m(t)}=\left\{\frac{x^0_l}{x_m}+\frac{E_m-W_m}{\beta_{ml}(W_{1l}-W'_m)}\right\}\cdot exp(k_l(W_{1l}-W'_m)t)$$

$$=\frac{E_m-W_m}{\beta_{mi}(W_{1i}-W'_m)} \tag{2-72}$$

$$\frac{x_{1i}(t)}{x_m(t)}=\frac{x^0_{1i}}{x^0_m}\cdot exp\{k_0(W_{1i}-W'_m)t\} \tag{2-73}$$

为了对单物种获得这种显式解，有必要作出假设：在竞争者过快地生长壮大起来以前，主宰拷贝和突变体之间已达到了"平衡"，以致于(2-64*a*)中的(小的)校正项中的比：

$$\sum_{k=1}^{\nu(\lambda-1)} x_{1k}/x_m$$

能够由(恒定)平衡比来代替。主宰拷贝的速率方程取得简单形式：

$$\dot{x}_m=k_0(W'_m-\bar{E})x_m \tag{2-74}$$

在此，W'_m与W_m仅有一项不同：只要$\sum_k x_{1k}$比x_m小[即$(1-L_m)\ll 1$]，这一项就很小。进一步的程序类似于情况*a*)。我们容易求解x_{1i}/x_m，若将

$$\frac{d}{dt}\left(\frac{x_{1i}}{x_m}\right)=\frac{\dot{x}_{1i}x_m-\dot{x}_m x_{1i}}{x^2_m} \tag{2-75}$$

插入(2-64)式。同理可求x_{ji}/x_m。该比率常常用于表达采取(2-51)式形式的$\bar{E}$，$\bar{E}$在插入(2-74)式以后获得了关于x_m的伯努利型微分方程。求这个积分类似于情况*a*)。

到此为止，我们的讨论限制在自我选择反应系统的理论特征上，而没有讨论任何现实的应用。对于那些理论特征，借助于计算机总可获得满意的解。上述考虑向我们表明，在什么范围我们可运用较简单的近似。事实上，仅仅只有不多的几个反应系统可以应用简单形式的线性一次速率方程。一般地，我们必须要考虑各种相互作用，首先是要考虑"信息"分子和"功能"分子之间的相互作用，而且还要考虑各类中的个体成员之间的相互作用。这可以导致整体的反应循环或反应网络，其中包括每一个体反应作用物的非线性速率方程，例如象米凯利斯—门顿型或更复杂的方程。第4到第6章将讨论这些更现实的系统，

并把它们与实验结果联系起来。我们将会看见，在本章所描绘的理论选择行为，在这些更复杂的系统中明确地再现出来，尽管它们的显式解也许会显示出重要的质的差别——诸如各种形式的周期性以及提供非常明确的选择奇点。我们还将看到，在作出关于自组织“活”系统起源的结论时，这些质的差别将表现出十分重要的意义。

3 选择的随机处理

3.1 决定论选择理论的局限

我们到此为止，是把选择作为决定论过程加以处理的。唯象方程清楚地说明了，在给定系统中哪一种拷贝将被选择。具选择优势（$W_i > \bar{E}$）的突变体一旦出现，其生长将不可避免地超出先前的分布。

这种关于选择的决定论描述，有两个重要的限制：

1．导致特定突变体的基本过程，本质上是非决定论的。自催化放大导致了“不确定”微观事件的宏观图式。①

2．生长过程自身也是统计涨落的过程。由于生长始于单拷贝，所以就必须考虑这种涨落，它们可能大大地更动决定论理论的结果。只有在有关大量物种的平均值情况下，决定论理论才有效。

与其平衡态比较，某些定态是亚稳的，这个事实又将带来一个困难。这些定态不可能使自己稳定，如果它们要长时间保

① 就我所知，约尔丹最先注意到受量子力学不确定性支配的基本事件的“放大作用”。

持，就需要调节。正是由于所有这些事实，使我们不得不从概率论的观点重新审度选择问题。大家将会发现，通过随机处理①，将导致对(决定论的)唯象理论作出重要修订。

3.2 平衡态附近的涨落

为了标征定态附近的涨落与稳定平衡态附近的涨落之间的差异，我们从经典的平衡涨落例子，即埃伦菲斯特缸子模型(参见图3)入手来开始讨论这个问题。

给定两个缸子以及(大量的)球，譬如说 $2N$ 个球，这些球任意地分布在两个缸子中。对这些球标以数字1到 $2N$。

"博奕"按如下方式进行：通过诸如抽签、掷骰子或能想象出的任何其他彩票机来随机地取出一个数，然后把这个数所对应的球从一个缸子转移到另一个缸子。如果不断反复这个程序，结果将是 $2N$ 个球平分在两个缸子中，而与初始分布无关。

这个模型是由 P. 埃伦菲斯特和 T. 埃伦菲斯特设想出来的，后来由特尔·哈尔和格林、卡克、克莱因及其他人进行了随机处理。科尔劳斯赫和薛定谔对此模型进行了实验检验。按照随机处理(例如，由卡克进行的)结果，此模型具有下述特征：

1. 尽管有涨落，该平衡仍然呈现出稳定分布。平均每个缸子中都将有 N 个球。

2. 在平衡态附近有涨落n，即一个缸子中有 $N+n$ 个球，而另一个缸子中有 $N-n$ 个球，n 可取从 $-N$ 到 $+N$ 的所

---

① "随机"理论是概率论向动力学问题的推广。这个词来源于 στοχαζομαι，即有目的、碰巧或猜测等意思。在《物理学百科全书》III/2 中，拉马克里舍南有极好的评述。

有值。类似于玻尔兹曼定律，这个模型能用分布函数来描述：

$$H=(N+n)ln(N+n)+(N-n)ln(N-n) \quad (3-1)$$

或者对于 $n \ll N$：

$$H=\frac{2n^2}{N}+\text{常量} \quad (3-2)$$

这表明，涨落是关于 $n=0$ 对称的。

3．在一个缸子中找到 $(N+n)$ 个球且在另一个缸子中找到 $(N-n)$ 个球的概率为：

$$P_0(n)=2^{-2N}\cdot\frac{(2N)!}{(N+n)!\,(N-n)!}\longrightarrow\frac{e^{-\left(\frac{n}{\sqrt{N}}\right)^2}}{\sqrt{\pi N}} \quad (3-3)$$

随机方程的这种概率是平稳的，即是与时间无关的，无论何时总有涨落 n_0。此概率分布是关于 $n=0$ 对称的(高斯型)，其中的半宽与 $\sqrt{N}$ 成比例。这与出现了 $n=N$ 的巨涨落截然不同。我们有：

$$\frac{P_0(n=N)}{P_0(n=0)}=2^{-2N}\sqrt{\pi N} \quad (3-4)$$

4．这个关系完全适用于 $\tau(n=0)/\tau(n=\pm N)$，即"再现时间"(即再出现相同宏观态的平均时间)的比。对于 $n=0$，这个时间取最小值。

这就得到如下重要结论：

平稳态是一种"稳定"态。而涨落则是*自我调节的*；它在一个方向上的偏差越大，使它反转即恢复平衡的概率就越大。平均涨落与 $\sqrt{N}$ 成比例，所以它们对于大的 N 就不重要了。由(3-4)式给出的再现时间比表明，如果 N 是一个大数，那么巨涨落实际上极其稀罕。对于与涨落比较，澄清不可逆过程的不同本性，该模型具有重要的历史意义。

3.3 定态的涨落

在前面的例子中，两个缸子出现的状态是等价的，2N个等价的球在这种状态中取得了一种稳定平衡态。现在改变一下我们的模型：不用两个缸子，只用一个缸子，每个球不是给予一根签，而是两根签，其中一根是白签，而另一根是黑签；一旦抽出一根白签，就把另一个球加入此缸中，而一旦抽到一根黑签，则从此缸子中移走一个球。实际上，考虑到这些球都是等价的，我们不需要计数。不过，如果我们的兴趣在于单物种的进化，我们可以保持标记数字的球；一旦我们抽到一根白签，则相应的某个数字就加倍。我们还必须保证在缸子中的每一个球都由两根签(一黑一白)代表，所以，移去或加入一个球总是要移去或加入两根相应的签。

这个模型代表了一种典型的定态问题，在此生成和分解的概率相等，都与所存在的小球数成比例。决定论方程将再次显示一种与时间无关的分布。不过，由于加入和移去的速率之涨落是互相独立的，因而不能自我调节，所以这种分布是亚稳的。这个例子与化学速率问题密切联系，在下面的讨论中，我们将用更实在的术语重新叙述这个例子。

我们可以回想起在2.1中引入的"信息匣"。该匣壁是半透的，高能态和低能态的单体字符都可以通过匣壁，而所有的大分子信息载体以及任何复制机构(酶等)都保持在匣内。匣中的条件有利于大分子合成；然而只能出现模板指令的合成即复制。让我们考虑3个问题，它们将对选择的随机方面有更多的阐明，在后面还将对它们进行定量处理。

1. 给定 N 个不同序列（为简单起见假定序列的长度都一

样)，N 是一个大数，每个序列只有一个拷贝；只有模板指导过程才能生成新序列，而且自体复制是精确的，即不会产生差错($L_k=1$)。所有生成和分解的速率因子 F_k 和 R_k 都是相同的，即 $W_i=E_i=W_k=E_k=0(i, k=1, 2, \ldots, N)$。(一致的 $W_i=E_i$ 归因于 $L_i=1$) 按照决定论，该系统应处于定态，其中不发生任何变化。对于任何特定的单拷贝，情况显然不可能是这样。对于整个集合，至少在某个时间间隔内，如果我们对不同的拷贝(它们的速率性质都是简并的)不加区别，那么情况可以近似如此。我们的问题是，对于信息载体的全体及其某一特定类，它们真正的命运是什么？(注意，我们从 N 种不同复本入手，其中的每一复本可以代表一条不同的"信息"。)

2．在此假定与例1的条件相同，但是现在允许有复制错误，即 $L_k<1$。可是，对所有拷贝仍要求 $W_i=W_k(i, k=1, 2, \ldots, N)$。现在，由于复制过程中的错误，可能生成新的序列。我们再一次要求，总信息量和个体信息量都与时间相关。

3．从例2的条件出发，我们最终允许生成和分解有不同的速率参数W_i、$W_k(i, k=1, 2, \ldots, N)$。若无流量控制，这个系统起初只能处于定态，此时平均总生成速率等于平均总分解速率。不过，人们可以按照本文第 2 章说明的约束条件来控制单体字符和(或者)溶剂的流入量，从而可以保持定态。我们在此特别感兴趣的是个体物种的进化行为，尤其是以具有选择优势 $W_i>\bar{E}$ 的单突变体拷贝出现的那些个体物种的进化行为。

着手数学处理前，我们尽量把所期望的行为条理化。在例1中，对于任何外加的新信息(它们只可能已由复制错误所引入)，这个系统显然是"封闭的"。另一方面，一旦某一个并非多余的信息载体在被复制之前分解了，就会丢失信息。这种情况将稳步地出现，因此个体的并非多余的信息总量将稳步地减少。

如果信息载体的总数非常之大，那么这种"负"涨落可以被其他拷贝的"正"涨落(即增生)所抵偿。这样一来，在整个群体消亡之前，个体的信息量将集中到几种(或甚至一种)相当多余的序列。因为生成涨落和分解涨落互相无关，所以整个群体必然有这种命运。如同图2-6所示(参见插图说明中的讨论)，甚至可能在相对短的时间内出现总灭绝。"相对短"，是与3.2中讨论的相应的涨落再现时间比较而言。只要对单信息载体不加区别，就完全可以预期第2个例子中的总群体命运亦是如此。

在此所有物种的速率相等，总生成正好由总分解抵消。可是，单信息载体(即个体信息)的进化行为与第一种情况显著不同。如果放大因子 A_k 正好等于分解因子 D_k，而品质因子小于1，那么因为对每一种物种 $W_k=A_kL_k-D_k<0$，所以每种物种都必将消失。在复制过程中，这个系统由借助于误差平稳地产生出的新信息而得到补偿。它"无规则地在信息空间漂移"，直到整个群体被"涨落突变"毁灭，这跟第1个例子的情况一致。

只有在第3个例子中，我们才可以期待稳定的、可复制的行为。系统在此将选择具有最大 W_k 的物种(假定任一物种都有 $W_k>0$)。只要这种物种仅以几个拷贝存在，它就仍然面临被涨落消灭的危险。可是，随着它的生长壮大，它就将变得更稳定；按照决定论的方程，它终将统辖整个群体。一个有趣的问题产生了：如果其概率有利于生长，图2-6所示的随机行走中是否有某个"不可返回点"？每个飞行员都知道，跑道上有这种点，一旦通过了这个点，就必须起飞。类似地，我们可以在浓度轴上找出这样一个"临界"点，突变体一旦达到了这个点，就不可能"返回"。

只有进行定量的理论处理，才能得到所有这些问题的定量答案。

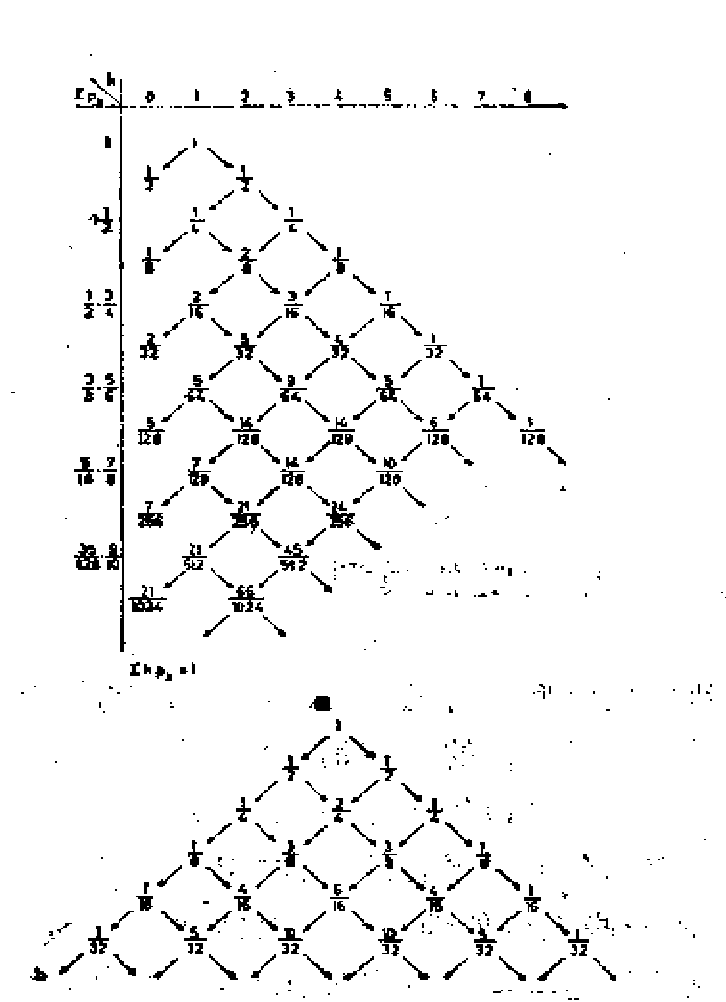


图 2-6 定态模型中的概率示意图 水平数字 k 标明每一拷贝的多余度。注意这个图只代表了概率而不代表时间进化。每一步代表了群体的一次变化，且增加和消去的机会相等。移动的时间间隔随 k 的增加而减小，即与 k 成反比。选择这种表示法，是为了揭示由 $k=0$ 的"灭绝"状态所带来的随机行走的不对称性。

如果没有这种不对称，即如果每一序列的状态追循前一序列的状态的概率为 1/2(对负的 k 亦如此)，那么"帕斯卡三角形"中的数，是二项式系数除以相应行的系数的和所得到的。正如所见，灭绝条件以不对称方式改变了这些概率，所以两个三角形只有右边界线相一致。

如果想要预见灭绝的概率，那么可从任一数字 k 出发，并考虑以出发点为顶点的三角形中的随机行走。因为每一方向的移动概率是相等的，所以达到灭绝状态(即 $k=0$)之前平均要行走 k^2 步。由于时间间隔的减少

与k成反比，所以我们能够预言，灭绝所需时间等于k和k^2之间的时间间隔乘以平均基本时间$1/F_i$。这个时间应该大于但接近于k/F_i。在$t=k^2/F_i$，几乎肯定发生灭绝。模型中（从k出发）达到零状态的最短时间由下式给出：

$$\frac{1}{F_i}\left(\sum_{i=1}^{k}\frac{1}{i}\right)\longrightarrow\frac{1}{F_i}(c+\ln k),\text{对于大的}k,$$

式中c是欧拉常数

$$c=-\int_0^\infty e^{-t}\ln t\ dt\approx 0.5772.$$

3.4 马尔科夫链的随机模型

对各种进化模型进行随机处理，这是当前的一个课题。在此，我们对于运用先前的某些简单线性模型获得的结果所引起的上述问题，多数可以给出原则性的回答。

巴塞洛梅不久前对平稳的线性“出生和死亡过程”进行了随机处理；其他一些人也处理了一些类似问题，例如，早在1940年德尔布吕克就讨论了简单的自催化（正）反应。在另外的地方，还能找到关于随机理论应用于化学速率过程的文献综述。

以下的讨论，以巴塞洛梅的精致处理为基础，他在处理中运用了杜布的Q矩阵方法。表2-9和表2-10总结了杜布方法和巴塞洛梅程序的基本原理。

运用决定论方程(2-10)式，以及条件$\varphi_{ij}=0$、$F_i=0$和$R_i=$常量，这个问题由一个平稳马尔科夫链来表示

$$\{x_t, 0\leqslant t<\infty\} \tag{3-5}$$

时间(t)是一个连续参数；x_t是一个随机群体变量，它涉及到一个许多可数状态S_0，S_1，…的离散序列。从一个状态(S_i)到另一个状态(S_j)变化由转移概率$p_{ij}(t)$描述。这正是一个系统在

$t=0$ 时从 S_i 态出发，在时刻 t 到达 S_j 态的概率。对 $p_{ij}(t)$ 的了解得以确定，群体从一定的起始条件 $t=0$ 出发并于时刻 t 到达某一状态的期望值和均方差。应用于线性出生和死亡过程的转移概率列在表2-10中，这是我们继续讨论的基础。

表 2-9　杜布用于随机过程的Q-矩阵方法（依据巴塞洛梅）

1．要求

把通过状态 S_i 到 $S_j(i,\ j=0,\ 1,\ 2,\ \dots)$ 的转移概率函数 $p_{ij}(t)$ 定义为条件概率

$$p\{x_{t_0+t}=S_j|x_{t_0}=S_i\} \tag{1}$$

随机变量 x_t 将取值 S_j，如果它在 t_0 时具有值 S_i 的话。过程 $\{x_t, 0\leqslant t<\infty\}$ 可以是一个平稳马尔科夫链，只要转移概率满足如下条件：

$$p_{ij}(t)\geqslant 0 \qquad (i,\ j=0,\ 1,\ 2,\ \dots) \tag{2}$$

$$\sum_j p_{ij}(t)=1 \qquad (i=0,\ 1,\ 2,\ \dots) \tag{3}$$

$$\sum_j p_{ij}(s)p_{jk}(t)=P_{ik}(s+t) \qquad (i,\ k=0,\ 1,\ 2,\ \dots) \tag{4}$$

$$\lim_{t\to 0} p_{ij}(t)=\delta_{ij}=\begin{cases}1, 当\ i=j\\ 0, 当\ i\neq j\end{cases} \tag{5}$$

2．矩阵 $Q=(q_{ij})$ 的定义

$$q_{ii}=\lim_{t\to 0}\frac{p_{ii}(t)-1}{t}\equiv\frac{dp_{ii}}{dt} \qquad (t=0) \tag{6}$$

$$q_{ij}=\lim_{t\to 0}\frac{p_{ij}(t)}{t}\equiv\frac{dp_{ij}}{dt} \qquad (t=0) \tag{7}$$

把要求1和定义2运用于构造下面的微分方程系统($\dot{p}=dp/dt$)。

3．a) 前向系统

$$\dot{p}_{ik}(t)=q_{kk}p_{ik}(t)+\sum_{j\neq k} q_{jk}p_{ij}(t)] \tag{8}$$

b) 后向系统

$$\dot{p}_{ik}(t)=q_{ii}p_{ik}(t)+\sum_{j\neq i} q_{ij}p_{jk}(t) \tag{9}$$

前向系统描述转移前的最后一个时间间隔($t\to 0$)中发生的东西，而后向系统描述转移后的第一个时间间隔中发生的东西。它是这样表达的：对于前向系统是变动的终态之和，而后向系统是变动的始态之和。

4．一般解

若状态数有限，起始条件已给定［例如条件(5)］，则如杜布表明的，两个系统的唯一解可以用矩阵形式给出：

$$\mathbf{P}(t)=e^{t\mathbf{Q}} \tag{10}$$

$\mathbf{P}=(p_{ij})$及$\mathbf{Q}=(q_{ij})$均为矩阵，式中$e^{t\mathbf{Q}}$是指数级数展开式的逐项和：

$$1+t\mathbf{Q}+t^2\mathbf{Q}^2/2!+\cdots \tag{11}$$

在确定$p_{ij}(t)$及相应的q_{ij}时，该决定论方程的一些性质已给予了概率解释(参见表2-10)。

表2-10　马尔科夫链的线性出生和死亡过程

给出决定论方程

$$\dot{x}=(F-R)x \tag{1}$$

其解为

$$x(t)=x(t=0)e^{(F-R)t} \tag{2}$$

式中F和R是常量。

在t_1和$t_1+\Delta t_1$之间的小的时间间隔中，群体中净变化Δx_1为

$$\frac{\Delta x_1}{x_1}=F\Delta t_1-R\Delta t_1+O(\Delta t_1) \tag{3}$$

式中$O(\Delta t_1)$包括所有的高阶无限小量。选择Δt_1充分小，以保证在时间间隔中只有一次生成或分解的有限概率，这由下式给出：

$$Fx_1\Delta t_1+O(\Delta t_1)\ \text{或}\ R_{x1}\Delta t_1+O(\Delta t_1) \tag{4}$$

于是群体中只有加入或减去一个单个物种的变化，于是，只有$S_i\to S_{i-1}$或$S_i\to S_{i+1}$的转移是允许的，并且出现这样的转移只能通过一件基本事件（抵消掉±1的净变化的复合出生和死亡事件被排除在外）。而且，因为当系统达到S_0态时就“消亡”了，所以不难理解$S_0\to S_1$的转移概率为零。从如下展开式入手：

$$p_{ij}(\Delta t_1)=p_{ij}(0)+\dot{p}_{ij}(0)\Delta t_1+\ddot{p}_{ij}(0)\frac{(\Delta t_1)^2}{2}+\cdots$$

便可按表2-9中的(6)和(7)式构造出$\mathbf{Q}$矩阵，再与本表中的(4)式比较便可获得

$$p_{i,i-1}(\Delta t_1)=q_{i,i-1}\Delta t_1+O(\Delta t_1); \tag{5}$$

$$p_{i,i+1}(\Delta t_1)=q_{i,i+1}\Delta t_1+O(\Delta t_1); \tag{6}$$

$$q_{i,i-1}=iR;\ q_{i,i+1}=iF;\ q_{ii}=-i(F+R) \tag{7}$$

($i=0, 1, 2, \cdots$，代表群体的状态)

于是又能够构造表2-9中说明的“前向”和“后向”方程：

$$\dot{p}_{ik}(t)=-k(F+R)p_{ik}(t)+(k-1)Fp_{i,k-1}(t)+(k+1)Rp_{i,k+1}(t) \tag{8}$$

$$\dot{p}_{ik}(t)=-i(F+R)p_{ik}(t)+iFp_{i+1,k}(t)+iRp_{i-1,k}(t) \tag{9}$$

这些方程可由"概率发生函数"方法求解，即运用如下定义的函数来求解：

$$\phi_i(s, t)=\sum s^k p_{ik}(t) \tag{10}$$

从 $\partial\phi/\partial s$ 和 $\partial\phi/\partial t$ 得到

$$(\partial\phi_i/\partial t)-(s-1)(Fs-R)(\partial\phi_i/\partial s)=0 \tag{11}$$

借助辅助方程

$$\frac{dt}{-1}=\frac{ds}{(s-1)(Fs-R)} \tag{12}$$

在 $F\neq R$ 和 $F=R$ 两种情况下分别进行积分，导出了概率发生函数的偏微分方程的一般解：

1．$F\neq R$：

$$\phi_i(s, t)=\left[\frac{(Re^{(F-R)t}-R)-(Re^{(F-R)t}-F)s}{(Fe^{(F-R)t}-R)-(Fe^{(F-R)t}-F)s}\right]^i \tag{13}$$

以 s 的幂级数展开导致了 s^k 的系数值，将其与(10)式比较则得到概率

$$p_{ik}(t)=\sum_{n=0}^{k\text{或}i}(-1)^n\binom{i}{n}\binom{i+k-n-1}{k-n}F^{(k-n)}R^{(i-n)}\cdot[e^{(F-R)t}-1]^{i+k-2n}$$

$$\cdot[Fe^{(F-R)t}-R]^{-i-k+n}[Re^{(F-R)t}-F]^n \tag{14}$$

若 $0<k<i$，则上和限为 k；若 $k\geqslant i$，则上和限为 i。我们有

$$\lim_{t\to 0} p_{ik}(t)=\begin{cases}1 & (k=i)\\ 0 & (k\neq i)\end{cases} \tag{15}$$

i 的期望值由下式给出

$$\varepsilon_i(t)=\left(\frac{\partial\phi_i}{\partial s}\bigg|_{S=1}\right)=i\cdot e^{(F-R)t} \tag{16}$$

这即决定论的结果。

均方差为

$$\sigma_i^2(t)=\left(\frac{\partial^2\phi_i}{\partial s^2}\bigg|_{S=1}\right)+\varepsilon_i(t)-\varepsilon_i^2(t)$$

$$=i\frac{F+R}{F-R}e^{(F-R)t}[e^{(F-R)t}-1] \tag{17}$$

2．$F\neq R$：我们类似地获得

$$\phi_i(s, t)=\left[\frac{Ft-(Ft-1)s}{(1+Ft)-Fts}\right]^i, \tag{18}$$

$$p_{ik}(t)=\sum_{n=0}^{k\text{或}i}(-1)^n\binom{i}{n}\binom{i+k-n-1}{k-n}$$

$$\cdot(Ft)^{i+k-2n}(Ft-1)^n(Ft+1)^{-i-k+n} \tag{19}$$

$$\varepsilon_i(t) = \left(\frac{\partial \phi_i}{\partial s}\bigg|_{S=1}\right) = i; \tag{20}$$

$$\left(\frac{\partial^2 \phi_i}{\partial s^2}\bigg|_{S=1}\right) = i(2Ft + i - 1)。 \tag{21}$$

3.5 三种选择原型的定量讨论

例1. 让我们回到3.3中引入的三个定态问题中的第一个问题，并看一看由如下参数所说明的系统的生存和灭绝的概率是多少：

$F_k = F = R_k = R$（对任何物种 k 的出现）。

$t=0$ 时从 N 种物种的总群体出发。从表2-10的(19)式得到灭绝的概率（对 $F=R$）：

$$p_{NO}(t) = \left[\frac{Ft}{1+Ft}\right]^N \tag{3-6}$$

$$\longrightarrow e^{-N/Ft}, 若 F_t \gg 1 \tag{3-7}$$

对 $Ft \gg N$，则这个概率趋近1。在 $Ft=N$ 时，我们已有

$$p_{NO}(t=N/F) = 1/e \tag{3-8}$$

即，群体灭绝的概率大于1/3。在 $Ft=N^2$，有

$$p_{NO}(t=N^2/F) = 1 - 1/N \tag{3-9}$$

则灭绝几乎是肯定的。

群体的期望值仍然等于 N，且与时间 t 无关[参见表2-10中的(20)式]。不过，均方差随时间增加。在 $t=N/F$ 时，它变成

$$\sigma_N^2(t=N/F) = 2N^2 \tag{3-10}$$

从而正如在图2-6的讨论中提示的，我们发现，$t=0$ 时由 N 种物种构成的定常群体，一般在时间 N/F 到 N^2/F 之间将存活下

去；在 $t=N^2/F$ 时几乎肯定要灭绝。

我们现在要考虑 N 种不同的个体物种的命运，其中每一物种在开始时都只有一个拷贝。

按表2-10中的(19)式，倍增 k 倍的单物种的概率为：

$$p_{1k}(t)=\frac{1}{(Ft)^2}\cdot\frac{1}{(1+1/Ft)^{k+1}}$$

$$=\frac{1}{Ft(1+Ft)}\left[\frac{Ft}{1+Ft}\right]^k\text{（对 }k\geqslant 1\text{）}\tag{3-11}$$

或者，它消亡的概率：

$$p_{10}(t)=\frac{Ft}{1+Ft}\tag{3-12}$$

对所有 $k\geqslant 1$ 的(3-11)式求和，并结合(3-12)式得出

$$\sum_{k=0}^{\infty}p_{1k}(t)=\frac{1}{1+Ft}+\frac{Ft}{1+Ft}=1\tag{3-13}$$

与表2-9中的(3)式一致。

对任何 $k\geqslant 1$，在

$$Ft=\frac{k-1}{2}\tag{3-14}$$

处，$p_{1k}(t)$ 作为 t 的函数将通过极大值；极大值为

$$(p_{1k})_{max}=\left(\frac{2}{ek}\right)^2\tag{3-15}$$

（e 是自然对数的底）。例如，如果我们要问，在什么时刻对哪一个 k 获得的极大值 p_{1k} 等于 $1/N$，那么结果是

$$k=\frac{2}{e}\sqrt{N}\tag{3-16}$$

或者

$$t=\frac{1}{F}\left(\frac{\sqrt{N}}{e}-\frac{1}{2}\right)\tag{3-17}$$

让我们把 $p_{1k}(t)$ 看作 k 的函数。对于 $Ft \gg 1$，我们获得渐近结果：

$$p_{1k}(t) \longrightarrow \frac{e^{-k/Ft}}{(Ft)^2} (k \geqslant 1) \tag{3-18}$$

这种分布即在给定时刻 t_1 把 $p_{1k}(t)$ 看作 k 的函数（k 取对数标度），$t_1 = N/F$ 时如图 2-7 所示。

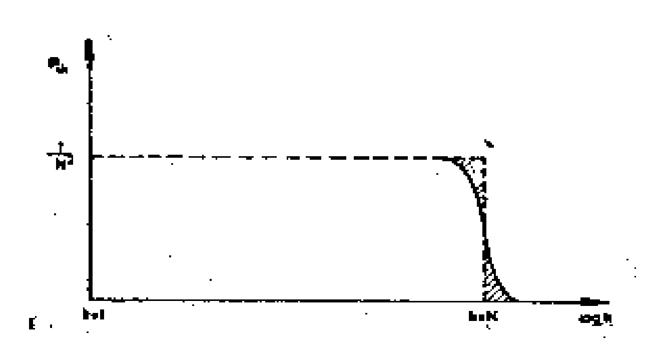


图2-7 按(3-18)式，$t_1 = N/F$ 时的"生存者"的概率分布。

正如所见，在 k 达到 N 的数量级时此概率将指数地减小，在此之前它几乎与 k 〔$p_{1k}(N/F) = 1/N^2$〕无关。同时，灭绝概率为：

$$p_{10} = 1 - 1/N \tag{3-19}$$

式中

$$\int_0^\infty e^{-k/N} dk = N \tag{3-20}$$

在 $k = 1$ 至 N 的范围，我们可近似地把概率分布看作恒量 $1/N^2$；$k = 0$ 时跃迁到 $1 - 1/N$。这样一来，在 $t = 0$ 时从某个拷贝出发，在1到 N 之间，k 被放大到任一程度的机会近似相等。因为在起点($t = 0$)我们有 N 种不同的单拷贝，所以 $t = N/F$ 时其中一种拷贝被放大到某个大于 $N/2$ 的数的机会已是 50%（即 $N \sum_{k=N/2}^{\infty} p_{1k} = 0.5$）。从 $p_{10} = 1 - 1/N$ 可知，在此同时，大多数其他拷贝都已

死亡。

每一物种的期望值 $e=\sum_{k=0}^{\infty}kp_k$ [p_k 依据 (3-11) 式] 仍然是"1"，且与时间无关；而对于每一物种的增加(例如 $Ft=N$)的均方差为：

$$\sigma_1^2(\cdot t=N/F)=2N \tag{3-21}$$

我们的结论如下：

在总信息量消亡之前，个体信息量集中到了几种甚至一种高度多余信息的载体上。这种由生成过程的自催化本性引起的选择行为，代表了典型的"生存者生存"情况。一个群体，若其中每一个体都具有相等的生存机会，则对于选择，除了结果即生存的事实以外，决无其他标准。

处于定态时，这种类型的纯粹"随机"生存是不现实的，因为品质因子决不可能正好等于1。它应与达尔文型的选择相区别，因为我们能够提出达尔文选择的最优生存标准。

例2． 在2.3中举出的第二个例子中，我们必须定量地区别总群体的解与单一的个体解。对于总群体，我们可使 $F=R$(未区别任何物种)。因此，总群体的行为将与上述的行为完全一致，即在时间超过了 N^2/F 的数量级时完全灭绝。

然而，对任一单物种 k，则因为 $L_k<1$ 而有 $F_R<R_R$，在此假定所有的个体物种都是关于 F_R、R_R 或 L_R 简并的。于是，表2-10中(14)式表明，对 $t\gg1/(R_R-F_R)$ 有 p_{10} 趋近 1，即

$$p_{10}(t)\to1-(1-F_R/R_R)e^{-(R_R-F_R)t} \tag{3-22}$$

对每一拷贝从1衰退为0的期望值为

$$e_i(t)=e^{-(R_R-F_R)t} \tag{3-23}$$

并且均方差趋近于：

$$\sigma_1^2(t) \longrightarrow \frac{R_R + F_R}{R_R - F_R} e^{-(R_R - F_R)t} \tag{3-24}$$

每一个物种都将消亡。与第一个例子中生存者的寿命相比较，任一物种都决不会有更长的寿命。而由于产生误差，信息量便会平稳漂移。该系统在其寿命期间即在小于 N^2/F 的时间内，但时间又适当超过$\frac{1}{R_R - F_R}$时，会扫掠过大量信息而不稳定地复制任何给定拷贝。可是正如前面已提到过的，这个集合作为一个整体，其命运类似于例 1 中的情况。

例 3. 只有存在有限变化的速率参数，才可能发现稳定的、可预见的选择。不过，即使在此种情况下，选择仍然有一定程度的"不确定性"。首先，对$F>R$，我们得到生长的群体期望值：

$$\varepsilon_i(t) = ie^{(F-R)t} \quad [参见 (2\text{-}23) 式]$$

它与决定论的理论一致。其次，在$t=0$时，对于大的 i，涨落处于预期的$\sqrt{i}$因子之内（i 是起始值）。现在，让我们观察具有 $W_m=\bar{E}$ 的某个已平衡的分布。那些具有有限"选择优势" $W_{m+1}>W_m$ 的物种，有一定的机会被选择，但决不是一定被选择。按照表 2-10 中的(14)式，如果 $t=0$时存在k种拷贝，该物种消亡的机会是：

$$p_{k0}(t) = \left\{1 - \frac{F_m - R_m}{F_m - R_m exp[(R_m - F_m)t]}\right\}^k \tag{3-25}$$

它在 $t\to\infty$时趋近于

$$\lim_{t\to\infty} p_{k0}(t) = \left(\frac{R_m}{F_m}\right)^k \tag{3-26}$$

这个结果中，包括了对3.3中的最后一个问题的准确回答。如果具有选择优势 $W_{m+1}>W_m$的某突变体$(m+1)$出现于先前已平衡的群体中(即 $W_m=\bar{E}=0$)，那么原先存在的这个拷贝仍然最终

可能消亡，其最终消亡的概率为 R_m/F_m，而它的生存机会是 $1-R_m/F_m$，生存机会将随着突变体拷贝的冗余度（k）的增加而增加。可是，线性模型中不存在真正的"不可返回点"。按(3-26)式，对 $F_{m+1}>R_{m+1}$，灭绝概率将随 k 的增加而稳步地减小，对大的 k 只以渐近方式达到零。

不过，我们可定义"半返回点" k_t，这是一个概率的阈值：

$$\left(\frac{k_m}{F_m}\right)^{k_t}=0.5 \tag{3-27}$$

或定义相应的"弛豫"点 $k_{1/e}$

$$k_{1/e}=\frac{1}{ln(F_m/R_m)} \tag{3-28}$$

对 $(F_m-R_m)\ll R_m$ 有

$$k_{1/e}=\frac{R_m}{F_m-R_m} \tag{3-29}$$

在这种条件下，$\lim\limits_{t\to\infty} p_{k0}$ 可以表达为

$$\left(\frac{R_m}{F_m}\right)^k \longrightarrow e^{-k\frac{F_m-R_m}{R_m}} \tag{3-30}$$

对各种 F_m/R_m，表 2-11 给出了相应的一些弛豫点 $k_{1/e}$ 的值。由此可见，小的选择优势几乎没有生存和取得支配地位的机会，这正如决定论的理论所预言的。这个事实进一步突出了选择过程的不确定本性。如果能够给出一个特定突变体出现速率的随机表达式，那么这个表达式必须乘以 $\left(1-\lim\limits_{t\to\infty} p_{10}\right)$，以便得到其宏观出现概率的表达式。

表 2-11 具有选择优势 $F_i>R_i$ 的突变体生存的随机阈值

F_i/R_i	$(F_i-R_i)/R_i$	$k_{1/e}$
2	1	1.44
1.3	0.3	3.82
1.1	0.1	10.5
1.01	10^{-2}	10^{2}
1.00001	10^{-5}	10^{5}

刚才所作的结论限制在线性生长系统。这些计算结果既被推广到由引入恒力或恒流条件的系统中，也被推广到(真正的)非线性生长系统中。非线性系统对于活系统的“集聚”具有特殊的意义(参见第6章)。

这种随机处理本质上以巴塞洛梅的线性的出生和死亡模型为基础，并已推广到由选择标准指示的真正定态，它对决定论的唯象进化理论提供了一些重要的修订。它不仅强调基本过程的不确定本性，而且相当清楚地揭示了，一些从决定论的理论导出的说法，若要能正确描述进化过程的本质特征，就必须先作出修改。

4 基于互补识别的自组织：核酸

4.1 真实的“自我指令”

到此为止所详加讨论的选择理论，虽然具有很一般的本性，但仅限在简单的准线性系统。“准线性”指的是由(2–32)式所描述的任何系统，即逆流项可略而不计，它具有如下简单形式：

$$\dot{x}=k_0(W_i^0-\bar{E})x_i \tag{4–1}$$

其中“选择价值”W_i由一常量所表示。我们应注意，无论是未约束生长的原始速率方程，还是由加上选择约束后所形成的速率方程，都不可能是真正的线性微分方程。原始方程中的“生成”项包含了“计量函数”$f_i(m_1,\cdots,m_\lambda)$，即高能单体的(在一般变量中的)浓度函数，它的精确形式有赖于模板指令聚合过程的机制，这个过程要加上(自催化)项x_i。只有按照恒总组织条件缓冲高能单体，才能在简单的“自我指令”的过程中把选择价值看作恒量。不过，在相同的条件下，我们在微分方程中引入了一个与浓度相关的(即含x_i的)函数$\bar{E}$。所以，“准线性”只可能涉及$W_i^0x_i$项，此项意指W_i^0不进一步依赖于x_i或任何x_k(它作为速率方程系统中的一个变量)。

在这些条件下，(4-1)式描述最简单的真正"自我指令"情况，此时特定序列i的生成指令由模板i自身给出。

我们如何能想象自然界中出现作为一种普遍现象的这种类型的自我指令呢？我们当然知道，许多特殊的自催化过程中一定的反应产物对自身的生成具有反馈。不过我们在此要求更高一些：聚合过程的任何产物即任何特定的序列，应该指令它自己的复制品的生成。

作为一个简单的例子，我们可以考虑在低pH值时生成聚腺嘌呤核苷酸(聚-r-A)。已知在pH<4时，聚-r-A通过质子化腺嘌呤残基的特定配对，从而形成双螺旋结构。它与沃森和克里克的结构不一样，这种双螺旋包含了平行取向的结合(即两链的走向都是3′⟶5′)。而且，碱基质子化作用本质上是中和骨架上磷酸基团的负电荷，从而使得这种结构在低的离子强度时更稳定。另一方面，这种双螺旋的行为又类似于沃森—克里克结构。波尔施克在我们实验室中进行了详细的热力学和动力学研究，从中可以得出这个结论。从聚-r-C的观察亦能发现类似现象，其中的配对既需要质子化作用物又需非质子化作用物，所以双(以及三)螺旋的生成被限制在相当窄的pH范围内。

对其中的任何一种情况，人们的确能够想象某种"自我指令"模板，它指导复制出一致的复制品；象这种系统，可由(4-1)类型的具有恒定选择价值W_i^0的速率方程来描述。让我们对决定系统的选择和自组织行为的参数A_i、L_i和D_i进行一些更细致的分析。

放大因子A_i的准确形式依赖于模板指令的聚合机制。J·吉布斯以随机模型为基础，已讨论过几个机制。

最简单的模型，应该对聚合链包含的每一字符提供恒定的、一致的时间间隔。如果每条链的复制在新的复制开始之前已经

完成，则复制的时间常数必定表现出直接与链的长度相关。如果此过程发生在远离平衡态，它就与ν(链中的字符数)成比例；但如果过程达到了平衡，它就趋近于与ν^2相关。不论是在正向的反应还是逆向的反应，在此概率都变得相等，反应于是类似于一个简单的线性扩散过程(传播时间在此与距离的平方成比例)。在目前的应用中，我们先不考虑系统接近平衡的情况，从而把比线形更强的那些与链长度的相关性排除在一边。

另一方面，有些机制产生一种可估计的比较弱的链长相关性：

a)此种过程可以是协同过程，因此集结需要一定时间(和长度)，这种时间可以适当大于"传播"的时间常量。因此，到达一定的"协同长度"后，复制时间便由集结时间所决定，从而与长度无关。例如，在生成低聚-*r*-A螺旋的碱基配对中已发现这种协同行为，在此"集结长度"包含了3个碱基对，"协同长度"约包含30个碱基对(在室温条件下)。

b)复制体复制完毕之前，在模板和复制体两者之间已出现了新链。对于酶的单链阅读(例如，由核糖体或RNA聚合酶进行的阅读)，这种现象是大家熟悉的，我们完全可以期望非酶的单链复制中有此种现象。正如 J· 吉布斯已经表明的，在这种复合的复制过程中，可以出现众所周知的"交通阻塞"问题。这与情况*a*)相反，这里仅仅发现相对短的序列出现链长相关，而链长到达一定长度后链长相关将消失。

如果把情况*a*)和情况*b*)结合起来，我们可以遇到相当弱的(弱于线性相关的)链长相关，这对于较短序列的复制就没多少有利之处了。然而，由于内部链折叠以及形成环，也许会出现模板效率极不相同的区域，所以不同长度序列的速率级数可以极不相同。而且，如果竞争中运用浓度不同的各种高能单体字

符，那么(通过A_i以及L_i)将大大有利于最丰富的字符，并可以导致相当一致的序列。如果涉及到酶①，则识别作用（类似于"集结作用"）可以包括模板序列的特定区域(很可能涉及到序列两端，这就如同$Q\beta$复制酶的情况，参见第7章)。

作为如下幂级数关系的结果，与放大因子相比，品质因子Q_{i0}对链长的限制更为强烈：

$$Q_{i0}=\prod_{k=1}^{\lambda} q_k^{\nu_{ik}} \tag{4-2}$$

式中下标k指的是不同的字符1，…，λ。如果单字符识别因子q_i是均一的，在表2-8中讨论的简单模型便能适用了(参见2.5)。

正如那里表明的，最大的可复制拷贝的字符数受到单字符识别的不确定性的限制，而且还(较弱地)依赖于速率系数的"扩散"。作为一个结果，单字符品质因子为0.99时(即误差率为1%)，生成可复制的序列将限制在数量级相当于字符数ν_i为100大小(如果"扩散"$A_i-\bar{A}_{k\neq i}$和$D_i-\bar{D}_{k\neq i}$趋近于零，那么甚至更小)。

如果识别是"平衡的"过程，那么单字符品质因子q可以用配对相互作用的自由能来表达。让我们考虑这种自我指令的字符复合步骤：

$$\boxed{i}+i_m \underset{k_D}{\overset{k_R}{\rightleftharpoons}} \boxed{i}\ldots i_m \overset{k_P}{\longrightarrow} \boxed{i}\ldots i_p \tag{4-3}$$

$\boxed{i}$代表模板字符，i_m代表高能单体，i_p代表其中的聚字符，k_R、k_D和k_P标志配对重组、配对解离以及字符复合(聚合作用)步骤的速率常数。"平衡的"识别意味着：

$$k_P \ll k_D \tag{4-4}$$

① 只要酶(例如某复制酶)代表了恒定的"环境因子"，即它不是"进化"系统的一部分，那么，酶促过程就可以包括在"准线性系统"的讨论中(参见第7章)。

下面叙述的测量表明，协同碱基配对(即某个"集结"区域的传播)的出现频率高达每秒 10^6 至10^7 次，K_D也可以大于每秒10^5次(GC 对)或 10^6 次(AU 对)。

假定"平衡识别"，我们可写出

$$q_i = \frac{(ii \text{ 配对的概率})}{(\text{所有 } i \text{ 的配对的概率和})} = \frac{m_i K_{ii}}{\sum_{k=1}^{\lambda} m_k K_{ik}} \qquad (4\text{-}5)$$

式中m_i或m_k是单体字符浓度，K_{ii}或K_{ik}是相应的(协同)配对的稳定常数。如果没有达到识别平衡，那么稳定常数可以由适宜的定态常数代替。稳定常数 K_{ik} 正好能够由生成配对的自由能来表达：

$$K_{ik} = exp(-\Delta G_{ik}/RT) \qquad (4\text{-}6)$$

只有所有单体字符的浓度被缓冲到具有相同值，m_k 才会下降，因而下述关系才保持有效：

$$1/q_i = \sum_{k=1}^{\lambda} exp[(\Delta G_{ii} - \Delta G_{ik})/RT] \qquad (4\text{-}7)$$

正如将要看到的，对任何无酶识别过程(相应于ΔG_{ik}中的差约为3千卡/摩)，很难产生出适当超过0.99的单字符q因子。

最后，对于现在的讨论来说，分解因子D_k不太重要。被选择物种的生成速率必须超过分解速率($W_i>0$)。就D因子而论，对于无酶过程，它们有与其他因子相类似的倾向，即它们也不利于长链。如果不形成保护性的大分子外壳，与短链相比，长链更容易水解。

对其他机制进行讨论，还可以得到一些进一步的结论。不过与下面的关于"自我指令"密码系统的进化的结论相比，它们就不重要了。

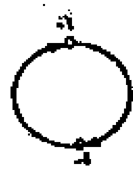
所有字符从一开始就以差不多的丰度出现是极不可能的。

另一方面，如果一种字符，例如一种腺嘌呤核苷酸，出现大量过剩，那么在竞争中就会如此强烈地有利于它的复合，以致主要生成十分均一的聚合序列(例如聚-r-A)。这种均一的序列毫无编码能力，因此，除了其他一些缺点之外，这种系统几乎对于任何继续进化的自组织都是不利的。

迄今我们还没有在自然界中发现此种简单的“自我指令”密码。原则上它可能已经存在过。但是，如果它曾存在过的话，那么在利用“互补指令”的系统方面，它处于很不利的地位。这样的一种系统，哪怕以一种占支配的高丰度字符出现，也会立即开始收集另一种互补的字符，从而构造出一个混合系统，这是密码发生的先决条件。

4.2 互补指令和选择(理论)

简单的(4-1)式不可能用于处理“互补指令”。这种复制过程可在“正”拷贝和“负”拷贝之间有某种选择，两种拷贝可分别记作$+i$(代表正链)和$-i$(代表负链)。收集($\pm i$)反应可用循环图表示：



每一次收集都由两个速率方程描述：

$$\left.\begin{aligned}\dot{x}_{+i}&=k_0(A_{+i}L_{+i}x_{-i}-D_{+i}x_{+i})\\ \dot{x}_{-i}&=k_0(A_{-i}L_{-i}x_{+i}-D_{-i}x_{-i})\end{aligned}\right\}\qquad(4\text{-}8)$$

其特征方程的解有两个本征值 λ：

$$\begin{vmatrix} -(D_{+i}+\lambda) & (A_{+i}\cdot L_{+i}) \\ (A_{-i}\cdot L_{-i}) & -(D_{-i}+\lambda) \end{vmatrix}=0 \tag{4-9}$$

$$\lambda_{1,2}=-\frac{D_{+i}+D_{-i}}{2}\pm\sqrt{4A_{+i}L_{+i}A_{-i}L_{-i}+(D_{+i}-D_{-i})^{2}} \tag{4-10}$$

这些本征值属于由浓度参数y_i表示的反应的某种“标准模”，y_i是x_{+i}与x_{-i}的线性组合。其中一个λ值总是负的。如果下式满足，则另一个λ值可为证：

$$A_{+i}L_{+i}A_{-i}L_{-i}>D_{+i}D_{-i}$$

x变量和y变量之间的关系可以写成向量形式：

$$\vec{y}_i=M_i\vec{x}_i;\quad \vec{x}_i=M_i^{-1}\vec{y}_i \tag{4-11}$$

式中矩阵M_i^{-1}是M_i的逆，M_i由本征向量的分量所决定。如果我们为简单起见，考虑(更为明显的)一致的分解项$D_{+i}=D_{-i}$（对多数处于恒力并调节了流量的实验，是一种良好的近似），则矩阵为

$$M_i=\begin{pmatrix} 1 & -\sqrt{\dfrac{A_{+i}L_{+i}}{A_{-i}L_{-i}}} \\ +\sqrt{\dfrac{A_{-i}L_{-i}}{A_{+i}L_{+i}}} & 1 \end{pmatrix};$$

$$M_i^{-1}=\begin{pmatrix} 1 & +\sqrt{\dfrac{A_{+i}L_{+i}}{A_{-i}L_{-i}}} \\ -\sqrt{\dfrac{A_{-i}L_{-i}}{A_{+i}L_{+i}}} & 1 \end{pmatrix}. \tag{4-12}$$

在同样的近似中，本征值成为

$$\lambda_{1,2}=\pm\sqrt{A_{+i}A_{-i}L_{+i}L_{-i}}-D_i \tag{4-13}$$

从物理学角度看，两个代表时间行为的实本征值规定一个循环。其中总是为负的一个本征值，描述了正链生成和负链生成之间的“平衡”弛豫过程：

$$y_{2i}(t)=y^0_i exp[-(\sqrt{A_{+i}A_{-i}L_{+i}L_{-i}}+D_i)k_0t] \tag{4-14}$$

它向某个 x_{+i} 与 x_{-i} 的恒定比衰退。如果平均生成项 $\sqrt{A_{+i}A_{-i}L_{+i}L_{-i}}$ 超过了分解项 $\sqrt{D_{+i}D_{-i}}=D_i$，[类似于 (4-1) 式中的正的 W_i^0]，另一个本征值就是正的。它代表着(±i)收集的自催化生长特质(对于 $D_{+i}=D_{-i}=D_i$)：

$$y_i^2(t)=y_{2i}^0 exp[(+\sqrt{A_{+i}A_{-i}L_{+i}L_{-i}}-D_i)k_0t] \tag{4-15}$$

对于选择来说，这是解的重要部分。矩阵 M_i 和 M_i^{-1} 允许我们把"标准"浓度变量转变成实际浓度量，且反之亦然，例如，对 $D_{+i}=D_{-i}$：

$$\begin{aligned}
y_{1i}&=x_{+i}-\sqrt{\frac{A_{+i}L_{+i}}{A_{-i}L_{-i}}}x_{-i};\\
x_{+i}&=\frac{1}{2}\left[y_{1i}+\sqrt{\frac{A_{+i}L_{+i}}{A_{-i}L_{-i}}}\,y_{2i}\right];\\
y_{2i}&=\sqrt{\frac{A_{-i}L_{-i}}{A_{+i}L_{+i}}}x_{+i}-x_{-i};\\
x_{-i}&=\frac{1}{2}\left[-\sqrt{\frac{A_{-i}L_{-i}}{A_{+i}L_{+i}}}y_{1i}+y_{2i}\right].
\end{aligned} \tag{4-16}$$

随着 $t\to\infty, y_{1i}\to 0, x_{+i}/x_{-i}$ 的"平衡"比为

$$\frac{\bar{x}_{+i}}{\bar{x}_{-i}}=\sqrt{\frac{A_{+i}L_{+i}}{A_{-i}L_{-i}}} \tag{4-17}$$

当正链与负链的生成速率相等，此比值等于1。如果从平衡比开始($t=0$)生长，则只有一个解 ($y_{1i}^0=0$)。从一种拷贝开始，即从正链开始 ($x_{-i}^0=0$)，是一种更共通的情况(参见第7章中的 $Q\beta$ 噬菌体实验)。那么，在达到恒定比之前，两个解对每一物种的时间变化都有贡献，且生长解占支配地位。

现在引入恒总组织的选择约束，对每一集合得到：

$$\begin{aligned}
\dot{x}_{+i}&=k_0\left[A_{+i}L_{+i}\frac{x_{-i}}{x_{+i}}+D_i)-\bar{E}\right]x_{+i}\\
\dot{x}_{-i}&=k_0\left[(A_{-i}L_{-i}\frac{x_{+i}}{x_{-i}}-D_{-i})-\bar{E}\right]x_{-i}
\end{aligned} \tag{4-18}$$

立即可见这些方程属于第二章讨论的一般类型，因为 x_{-i}/x_{+i} 或其逆不随 t 的变化而消失。

我们可把这个比 x_{+i}/x_{-i} 记作 z_i。立即可从(4—18)式中导出它的关于时间变化的微分方程：

$$\dot{z}_i=\frac{\dot{x}_{+i}x_{-i}+\dot{x}_{-i}x_{+i}}{x^2_{-i}}$$

$$=k_0[A_{+i}L_{+i}+(D_{-i}-D_{+i})z_i-A_{-i}L_{-i}z_i^2] \qquad (4—19)$$

积分得出有点冗长的表达式

$$z_i=\frac{\bar{z}_i(z_i{}^*+e^{-\alpha_i t})-2\beta_i e^{-\alpha_i t}}{z_i{}^*-e^{-\alpha_i t}} \qquad (4\text{-}20)$$

式中：

$$z_i{}^*=\frac{z_i^0+\bar{z}_i-2\beta_i}{z_i^0-\bar{z}_i};$$

$$\beta_i=\frac{D_--D_+}{2A_-L_-}$$

及 $\alpha_i=k_0\sqrt{4A_+A_-L_+L_-+(D_+-D_-)^2}$

它对于 $t=0$ 还原为起始比 z_i^0，而对 $t\to\infty$ 得到平衡比：

$$\bar{z}_i=\sqrt{\frac{A_+L_+}{A_-L_-}+\beta_i^2}+\beta_i \qquad (4\text{-}21)$$

把 $z_i(t)$ 代入(4-18)式，得到可以积分的非齐次微分方程。

互补指令系统的一般行为，类似于上述的“自我指令”系统的一般行为。每一次带信息的集合，现在由两种组分构成，从而能用二元列向量 $\bar{x}_i=\begin{pmatrix}x_{+i}\\x_{-i}\end{pmatrix}$ 来表示。先前的量 $W_i^0=A_iL_i-D_i$，现在可由下面的矩阵替代：

$$\begin{pmatrix}-D_{+i} & A_{+i}\cdot L_{+i}\\ A_{-i}\cdot L_{-i} & -D_{-i}\end{pmatrix}$$

这个矩阵有两个本征值，其中一个本征值代表($\pm i$)集合的平衡，而另一个描述集合的竞争生长性质。所以，后一个本征值及其相应的标准模 y_{zi} 进入了选择方程。在集合达到平衡以后，我

们可以简单地用和$x_{+i}+x_{-i}=y_i^*$代换标准模y_{2i}。代换式中 x_{+i} 和 x_{-i}都与y_{2i}成比例，于是可以在恒总力下把选择方程写成共同的形式：

$$\dot{y}_i^*=k_0(W_i^0-\bar{E})y_i^* \tag{4-22}$$

式中

$$W_i^0=\sqrt{A_{+i}L_{+i}A_{-i}L_{-i}}-D_i(\text{假定 } D_+=D_-=D_i)$$

以及

$$n\cdot\bar{E}=\sum_{k=\pm1}^{\pm N}E_k x_k;$$

$$E_k=(A_k-D_k)$$

其和遍取从$+k$到$-k$全体。

仍然假定$D_{+i}=D_{-i}$，由对$W_i^0=\bar{E}$所获的选择平衡得出

$$\frac{\bar{y}_{2i}}{n}\cdot\frac{(1+\sqrt{A_+A_-/L_+L_-})}{2}=\frac{y_i^*}{n}=\frac{W_i^0-\bar{E}_{k\neq i}}{\bar{E}_i-\bar{E}_{k\neq i}} \tag{4-23}$$

式中

$$\bar{E}_i=\frac{E_{+i}\sqrt{A_{+i}L_{+i}}+E_{-i}\sqrt{A_{-i}L_{-i}}}{\sqrt{A_{+i}L_{+i}}+\sqrt{A_{-i}L_{-i}}}=\frac{\bar{Z}_iE_{+i}+E_{-i}}{1+\bar{Z}_i}$$

($D_{+i}\neq D_{-i}$的表达式仍然有同样的一般形式，但是，W_i和E_i就必须依据(4-10)式及完全变换矩阵M_i和M^{-1}进行计算。)

可以发现，即使其中一个拷贝(例如$_{-i}$)其E值小于$\bar{E}_{k\neq i}$，只要满足下式就会出现选择：

$$W_i^0>\bar{E}_{k\neq i}。$$

不同的集合i_i再一次为选择而竞争。在不同的i_i中不存在任何其他耦合时，仅有某一集合i_m(与其误差的"慧尾"一道)，即有最高选择价值$W_m^0>0$的集合，将生存下去。此集合的选择价值包含了正链和负链的参数A_mD_m的几何平均值。这是一个有趣的结果，它表明两种链的复制参数同等重要。(注意，任何

算术平均值都将等价于“限速”项。)在眼下的情况中，速率较慢的组分向较高的定常浓度水平生长，从而达到速率较快的组分的速率。

“互补指令”的重要特征是，即便出现了一类字符大量过剩，系统总是必须积累至少两种不同的字符，这些不同的字符在被选择的物种中有几乎相等的丰度。由于错误复制，系统总要形成混合的序列。这是任何密码发生的前提，只要任一混合序列对它自己的复制表现出优势，密码就可以获得“意义”(即“代表着”有价值的信息)。正如后面将要看到的，刚开始时二元密码系统具有一定优势，此时不必预先转变成较高级的(例如四元的)密码形式就能提供优势。由于单(±i)系综的竞争本性，能够贮存起来的信息量限制在一类单载体〔即一个主宰拷贝及其(可复制的)错误拷贝“慧尾”〕或一个简并群的容量之内。因此，本质上由品质因子L_{+i}和L_{-i}所施予长度限制具有重要意义。关于核酸碱基互补识别的信息，我们可以从实验资料中获得。

4.3 互补碱基识别(实验数据)

4.3.1 单配对的形成

互补指令是以碱基A和U或者G和C之间分别形成独特的配对为基础的。在整个自然界，从噬菌体到人，这种独特的相互作用保证了遗传信息的精确阅读、翻译和放大。那么，这种相互作用的基础是什么呢？

一位生物化学家当然认为：“它们是特殊的酶”，而一位物理化学家也许倾向于认为：“它们是特殊的力”。两种回答都是对

的，因为我们已深入了解到一些这样的酶，也了解到存在一些特殊的相互作用，例如在不存在酶时的互补链的杂交作用。

图2-8示意了此种特殊的互补配对结构，这是克里克和沃森在他们的划时代论文中提出来的。可是，只要观察一下图2-9，立刻就会揭示出，此种氢键不足以解释自然界的明确选择。首先，不同配对的几何形状有差异（参见图2-9）。所以，两种沃森—克里克配对的同晶结构，对于形成包含所有四种核苷酸的均一的双链结构，特别对于共同聚合酶的进化适应，无疑地具有优势。另一方面，进化亦有可能始于例如A和U这样的两字符密码。而且，在酶专一适应之前，也许已经有了选择其他"互补性"的某些类型的密码子—反密码子相互作用。这些问题可通过适当的实验来回答。

图 2-8 互补碱基对

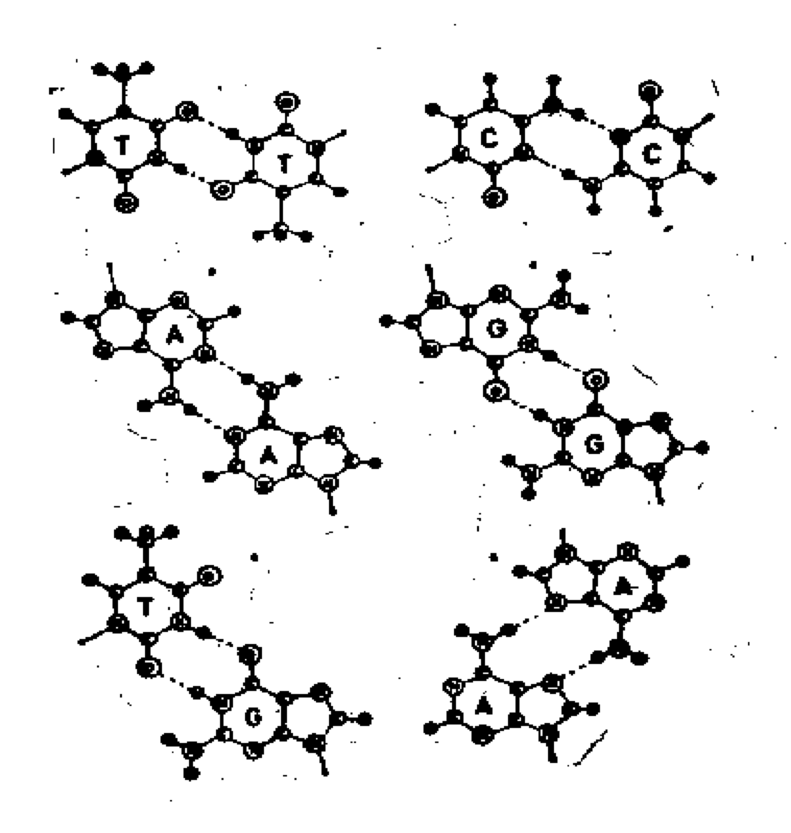


图 2-9 "非互补"碱基结合

已经作了一些实验。例如，对不同核酸碱基结合形成单配对的自由能，我们已可提出定量数据。曾有人试图用水介质作溶剂进行这种测量，结果是令人失望的：所有碱基对都不稳定。甚至在相当容易溶解的核苷酸的高浓度溶液中，也未检测出有一个单配对形成。我们已理解到结果为什么是这样，因为在诸如NH、NH_2、OH、CO:或N:这样的极性基因之间，唯一只有氢键才带来稳定化。于是，极性H_2O分子通过溶剂化与任何暴露的偶极性基团进行十分强烈的竞争。因此，进行相对稳定性实验时，使用非极性溶剂较为有利。

已运用多种非极性溶剂进行了测量。核苷酸可以溶解在这些溶剂中，测量是通过取代核糖上不同位置的非极性基团进行的（但不干扰碱基形成氢键的潜在能力）。麻省理工学院的里奇及其同事已做了大量此类工作，尤其是对各种已取代的核苷酸进行了红外研究。维也纳大学的库布勒和

伯克斯奇也报道了一些类似的研究。在我们的实验室中，芬克(部分得到霍普曼和埃格斯的合作)实现了关于形成配对的介电研究。他们还从弛豫测量中确定了动力学参数。

对所有这些结果有一种共同的看法。与可供选择的配对相比，互补配对AU和GC是最强的配对。表2-12中列出了一些稳定常数值，它们是由介电测量确定的。如图所示，*AU* 至少比AA或者UU都要至少稳定10倍，而GC在所有的G以及C的各种结合中是最有利的配对。GC对比*AU*稳定得多。由于G还表现出引人注目的“自配对”，因此不可能测定除了GC配对以外的G的任何其他异元配对的稳定常数。还应该注意，一些配对可以不同方式形成，表12中的平衡常数是总观值。

表 2-12 碱基配对稳定常数。25℃，$K_{ASS}[M^{-1}]$在非极性介质中的2′-3′-5′-0-取代核糖核苷。

C_6H_6 \ CCl_4	U	A	C	G
U	45 / 15	550	< 50	< 10^3
A	150	22 / 8	< 50	< 10^3
C	< 28	< 28	50 / 28	> 10^4
G	< 1.2·10^3	< 1.2·10^3	3·10^5	~10^4 / 1.2·10^3

所有配对的形成都是非常快的。速率常数表明，导致配对形成的每一次相遇以及此类配对的寿命一般都小于1微秒。

AU对和GC对明显相对稳定，这可能归因于一条强烈极化的氢键。对于AU和GC(但不是对于竞争者)，都只是XH…Y这

种类型的结合，即有几分酸性的NH基团(pK～9.5)与环氮(pK～4.5)之间形成氢键。对这条键的明显稳定性能够以量子力学为基础来理解。所以，在原则上，密码形成的重要前提可通过物理学理论加以解释。

总之，这些数据清楚地表明，在那些我们现在所谓的"互补的"碱基之间优先形成配对，"识别"是一个极快的过程。结果造成绝对稳定性相当低，以致难以在极性介质中检测出它们的形成。对于较长序列中阅读密码所需要的非常精确的识别，不可能从如此低的稳定性得到任何解释。为了更好地认识在序列的密码阅读中的碱基识别，必须研究发生在低聚或多聚物种中的协同作用。

4.3.2 低聚核苷酸和聚核苷酸中的协同作用

前面已经强调，在水解质中单碱基对极不稳定，难以用现在的技术检测出来。另一方面，众所周知，互补聚合链形成相当稳定的双螺旋结构，它只有在高温时才"熔解"。熔解曲线的形状表明，存在某种强烈的协同行为。协同性来源于相邻碱基对之间的一种较强的"堆积"相互作用，也来源于自由度"冻结"引起的关于形成螺旋的螯合效应。虽然这些协同效应大大增强了核酸的"指令"能力。

研究这种协同现象，一种简单方法是，从单个核苷酸出发逐步构造聚合体，研究随着链长增加时热力学和动力学性质的改变。对低聚物的构象变化，尤其是它们"全有或全无"类型的碱基配对行为，进行热力学和动力学分析，其重要优点是比较简单。对低聚物进行这种研究，困难之处在于为制备这种低聚物材料，需对多聚物进行降解、分离、收集和纯

化。在我们的(以及其他人的)实验室中①，几年来有各种小组从事着这项工作。

最初的动力学研究，是由波尔施克在做毕业论文时用低聚腺苷酸(链长为2～10字符）进行的。正如已提到过的，对于我们最初的研究，低聚-r-A和聚-r-A 在酸性范围（pH～4）形成双链螺旋结构，为我们提供了一个良好的模型。后来，波尔施克深入进行了这项工作，把它推广到低聚-r-A与低聚-r-U之间在中性pH条件时形成双或三螺旋（链长至多为18)。这些研究工作中，包括热力学研究(相图、由紫外光电仪观察得到的熔解曲线)与运用流和弛豫(温度跃迁)技术进行的动力学研究。

把这种工作推广到研究GC低聚物时，遇到了巨大的困难，原因在于这些东西聚集起来形成了更复杂的结构。波德尔在格丁根做博士后期间，已能研究下面的四聚物配对；

$$G_pG_pG_pC \qquad G_pG_pG_pU$$
$$C_pC_pC_pG \qquad C_pC_pC_pG$$

以及 G_pG 和 G_pG_pG 与聚-r-C 的配对。

马斯和一群来自格丁根和不伦瑞克—斯托克海姆的实验室的同事(里施勒、罗默和库茨)，与慕尼黑的扎考的实验室的一组人合作，研究了RNA(参见图 2-2) 的熔解行为和构象动力学，以及已知序列的片段的熔解行为和构象动力学，这些碎片是按照扎考描述的方法，对天然分子重复切割而获得的。

关于各种序列的特定配对，非常有价值的信息来自哈佛大学的多蒂的实验室，他们从事对低聚共聚物的研究，此类共聚物如：

$$A_pA_pA_pA_pU_pU_pU_pU$$
$$U_pU_pU_pU_pA_pA_pA_pA;$$

$$A_pA_pA_pG_pU_pU_pU$$
$$U_pU_pU_pC_pA_pA_pA;$$

---

① 鲍尔德温(斯坦福)，格罗什斯(耶尔)，戴维森(帕拉地那)，多蒂(哈佛)，弗里斯科(普林斯顿)，齐姆(拉·乔拉)。

$$A_pA_pA_pG_pC_pU_pU_pU$$
$$U_pU_pU_pC_pG_pA_pA_pA。$$

这些有价值的信息，还来自他们对各种与 tRNA 分子的暴露互补序列结合的低聚物研究（参见表 2-13)。多蒂的两位同事马丁和乌伦贝克在访问格丁根期间，与波尔施克合作，对上述低聚物进一步进行了热力学和动力学研究。所有这些研究的结果和(初步)结论，可以综述如下：

螺旋—卷曲平衡

形成双股螺旋时，"熔解"曲线表现出与链长有简单关系，即斜度随链长增加而增加。如果"熔解温度"（即从螺旋变成随机卷曲的半转变温度）的倒数对 $1/(\nu-1)$ （ν是双股螺旋结构中的最大碱基配对数的字符数）作图，则得到一条直线。该直线的斜度与$1/\Delta H$ 成比例；截距代表了$1/T\infty=\frac{\Delta S}{\Delta H}$，即"无限长"螺旋的熔解温度的倒数。"无限长"指的是比包含了30个碱基对的协同长度要长。因此，$T\infty$能够由较长链的聚核苷酸来确定。为获得这种结果，须对单链堆积进行校正，在不同的熔解温度(在能够由对单链聚合物进行实验来确定的范围内)，单链堆积的程度是不同的。短链从螺旋转变成随机卷曲可由"全有或全无"过程来表示。如果$\delta_i s$ 代表了与其中有 $(i-1)$ 个对的连续序列相邻的每一个碱基对的稳定常数，并注意到"集结参数"δ 在超过某个"集结长度"n 以后便向 1 趋近，那么我们可以把包含 $\nu>n$ 个碱基对的"全有或全无"转变的平衡常数写作

$$K_\nu=\bar{\sigma}s_\nu;\ \bar{\sigma}=\sigma_1,\sigma_2,\dots,\ \sigma_n \qquad (4\text{-}24)$$

动力学数据（参见后面）表明，在室温下，AU 螺旋的成核长度为 3，GC 螺旋的成核长度为 2，但温度大于 50℃ 时 GC 螺旋的成核长度也趋近于 3。这等于说，对 $n\geq3$，σ_n 趋向于接近 1。熔解曲线则表明，对$n\geq3$，随着每增加一碱基对，ΔH_s 或 ΔS_s 便有一恒定的(以及负的) 增量。按范霍夫关系式得到：

$$\ln s=-\frac{\Delta H_s}{RT}+\frac{\Delta S_s}{R} \qquad (4\text{-}25)$$

校正的$1/T_m$与$1/(\gamma-1)$成比例而不是与$\frac{1}{\gamma}$成比例，这一事实表明，当$n\leq3$时不可能出现恒定增量，特别是成核参数$\bar{\sigma}$跟温度的关系与s相反，所以$\bar{\sigma}s$几乎与温度无关（$\Delta H_{\sigma s}\sim0$）。虽然$\bar{\sigma}s$不是第一个碱基对的稳定常数σ_1s，但是从物理学角度看，ΔH_s基本上是来自堆积相互作用(注意，对每一单链，ΔH有相对高的堆积增量)，而不是来自互补碱基对之间形成氢键。所以，对第一个碱基对，ΔH值会相对低一些。由于与溶剂分子的竞争，这种配对是非常不稳定的(参见表2-18的常数值)。对较长的链，此种"全有或全无"模型不适用。系统的划分函数必须用满意的方程来评价，这正如阿普尔奎斯特和达姆尔对"交错拉链模型"的评价。波尔施克借助计算机程序，按照这种模型，已对实验数据进行了评价。

碱基配对的动力学和机制

借助动力学数据，就可能阐明协同碱基配对的机制。对短螺旋，"全有或全无"型转变已得到确认。只包含一个时间常量的弛豫谱，表现出一种二级的"全有或全无"机制，即两条(互补的)单链重组成具有最多碱基对的双螺旋。仅仅对于较长的链，才观察到附加的一级时间常量谱($t\ll1$微秒)，这代表了不固定的或交错的构象中的变化。二级过程的特征是有高重组速率，此速率标志一种成核阈值（k_R：$10^6\sim10^7$摩$^{-1}\cdot$秒$^{-1}$），然而它明显低于扩散控制的极限。对于所有大于3的链长（但低于协同长度)这种速率常数值几乎都相等。它们与温度的相关，虽说不大，但富有意义：对于AA螺旋（pH～4)和AU螺旋(中性范围，$T<50°C$)，以及较高温度下($T>50°C$)对于GC螺旋，情况都如此。这种相关由负的表观活化能来表示。

通过这些"活化"参数的绝对值，可以相当明确地规定成核长度。如果第一个碱基对的生成是限速步骤①，(碰撞）活化能应该为正[因为我们知道最初的配对不稳定（$\sigma_1s<1$）意味着配对的分解比生成更快]，我们便可预期有成核势垒。只有最初的配对代表了这种势垒，即如果一旦稳定双

① 仅仅观察到导致具最多碱基对的完整螺旋的过程。

配对核形成就有“拉链反应”的传播，那么速率常数 $k_R=\sigma_1 s k_{12}$，k_{12} 是生成第二个配对(即紧接第一个配对之后)的（一级）速率常数。由于 k_{12} 的活化能增量必为正，而且 σs 的 ΔH 增量几乎为零，$\sigma_1 s$ 的 ΔH 最多是一个小的负值，所以总表现活化能也在零附近，而决不可能接近于实验上的观察值−10千卡/摩（例如对于 *AU* 低聚物）。这个相对大的负值，仅仅与表达式 $k_R=\sigma s^2 k_{23}$ 一致，这个表达式描述生成第 3 个配对(k_{23})是限速步骤的过程。这样一来，对于 *AU* 序列，已配对的碱基三联体代表了稳定核。于是我们可把10^7～10^8秒$^{-1}$的数量级规定为传播的速率常数(k_{23})，以便与观察到的不固定构型的时间常量$\lesssim$1 微秒的弛豫谱一致。在室温下，GC 低聚物由于稳定常数“s”明显较高，所以能形成“双配对核”。它们的重组速率常数表现出正的活化能，这正是能量参数较高（与 *AU* 相比)所预期的。在熔解曲线上，它们都被(也是较高的)单链堆积参数“掩盖”了。

由单独测定分解速率常数，我们证实了这些结果。由于微观可逆性，我们须有：

$$k_D=\frac{k_{23}}{s\gamma^{-2}}(\text{因为 } k_R/k_D=\bar{\sigma}s\gamma) \qquad (4\text{-}26)$$

由于分母中的$s^{\gamma-2}$，所以活化能为正，其中包含了较大增量 $(\gamma-2)\Delta H_s$。

实验中的确观察到，这些预期值以及相应的 k_D 的绝对值随链长增长(以数量级计算)而减少。

4.3.3 关于识别的结论

表2-13中列出了从低聚核糖核苷酸和聚核糖核苷酸的平衡和速率研究中所获的数据。最吸引人的结果是形成三联体的倾向，不过它并不是必然的结果，即不是对20多种符号编码的前提，而是由于力学的巧合。少于三个字符的密码子将是非常不稳定的(至少对于A和U是如此)。多于三个字符的密码子，尤其是对于 G 和 C，又变得过于“粘性”。密码子—反密码子配对

表 2-13

碱基三联体和四联体（三核苷酸和四核苷酸）与 tRNA 的暴露区域（倾向于与反密码子）配对的稳定常数。数据取自多蒂等人的结果。$K_{ASS}[M^{-1}]$是在0℃、pH7且其中1.0mol/L NaCl、10mol/L $MgCl_2$ 和 10mol/L 磷酸盐的水溶液中测定的。K 值小于 400L/mol 时无法与"无结合"相区别。

甲酰甲硫氨酸 tRNA	K_{ASS}	酪氨酸 tRNA	K_{ASS}
AA |*UAC*| *UC*		*AA** |*AU***G*| *UC**	
AUG（正规的密码子）	1200 ±200	*UAC*（正规的密码子）	700
AUGA	13500	*UACA*	90000
AUGU	1400	*UAU*(3′-摆动)	700
AUGC	900	*UAUA*	37000
		苯丙氨酸 tRNA[a]	
		AY |*AAG**| *UC**	
AUGG	1000	*UUC*(正规的密码子)	900
GUG(5′-摆动)	1200	*UUCA*	10000
GUGA	9800	*UUU*(3′-摆动)	300
GUGU	1000	*UUUA*	1000

*A** = *N*(6) = 甲基-*A*；
*U** = 假*U*(ψ)；
*G** = 2-0-甲基-*G*；
*C** = 2-0-甲基-*C*。
a：与乌伦贝克的私人通信。

的寿命不会超过数毫秒，从而具有相应周转率的酶能最优地适应。在稳定性和速率之间的相同类型的优化，总是在酶—底物相互作用中发现。获得稳定性，意味着复合物的分解速率降低；为使它们不成为周转的限速步骤，它们必须与转变次数相

匹配。

还可以进一步注意到三联体的有趣特征。如果从链长大于$\nu=4$的低聚双螺旋出发推导互补的三联体的稳定常数，那么结果明显低于直接从碱基三联体或四联体测定的值（参见表2-14中由多蒂的实验室所测定的关于碱基三联体和四联体与tRNA的暴露区域相互作用的这种值）。对于始于二联体的GC配对的活化参数(参见正的值)，结果仍然如此。显然，由于碱基对自由排列成最有利的堆积交叠，所以短的暴露区域表现了较大的堆积相互作用。把这种区域扩展到扭曲的螺旋结构时，由于碱基配对受到更多的约束，则这种相互作用能将部分丢失。(类似的“空间”限制对“摆动”的GU对有效，而GU对仅在三联体的一端才有效。)

表 2-14 协同碱基配对的平均平衡参数和速率参数汇集

数据从水解质中的低聚核糖核苷酸（链长为3～18）的弛豫研究中获得。测量是由波尔施克和波德尔等人进行的，数据已外推到0℃：对于AU，pH=7，离子强度为0.05mol/L(卡可基钠盐)；对于GC，离子强度为0.1mol/L（磷酸盐缓冲溶液，pH=7.2)。

协同配对	σ [摩$^{-1}$]	s	ΔH^{a} [千卡/摩]	k_R [摩$^{-1}$·秒$^{-1}$]	$\Delta H^{\neq}(k_R)$ [千卡/摩]	k_{23} 或 k_{12} [秒$^{-1}$]
...A... ...U...	～10^{-3}	10	−11	10^6	−9	10^7
...G... ...C...	～10^{-3}	100～200	−15	10^6	+5～+7	10^7

a. 对“未堆积的”单链的外推值。(与堆积的单链有关的，实际测量值显著低一些)

注：因为精确值涉及到特定的配对结合(协同性)并强烈地依赖于实验条件（如离子强度等），所以表中只给出了速率及平衡常数的数量级。

不同区域的寿命，可借助于k_{23}(或k_{12})和s(或相应K_{12}的σs^{N})进行估计。

表 2-13 的数据涉及到主要在反密码子区域的配对，还可能涉及到其他暴露的区域。表 2-13 中的值高于表 2-14 中的值，这很可能是由于反密码子环的立体稳定化作用。因此，不可能从这些数据外推出 σ 值。与表 2-13 中端位错误配对情形中的，常数所表明的相比，在配对区域中的错误配对应产生更低的稳定常数（在噪声范围以内）。

如果我们运用表2-14中的数据，我们也许会得出关于q 值的某种结论，q值即决定互补指令精确度的单字符的品质因子。运用(4-5)式，对于互补相互作用例如对于AU(A 是模板元素)加以修改：

$$q_{AU}=\frac{m_U K_{AU}}{\sum\limits_{X=AUGC} m_X K_{AX}} \tag{4-27}$$

我们获得的值平均说来很难达到0.99，甚至对最适宜的条件亦如此。精确的测定将要求更多的关于在某互补区域的“误配”稳定常数的知识，以及各种最近相邻结合的稳定常数的知识。对AU而言，此值非常可能显著低于0.99；对GC而言，在特定条件下，它们也许要高一些，因为GC总是至少比AU稳定10～50倍(取决于最近的相邻)。这样的值代表了上限。它们要求互补识别的平衡先于字符复合成为聚合链，并且它们还与单体字符的浓度有关，在此所有单体字符已缓冲到浓度大致相等。在自然界中很可能找不到这些条件。我们可以认为，对于相对短的链，L值与1 相差较大。按照(2-45)式，对于密码载体的可复制形成，L必须保持在某个阈值之上。没有催化帮助时，具有特定碱基序列的核酸的这种可复制生成．只有对于相对短的链才是可能的，这些短链很可能不超过30～100 个字符（参见表 2-8，对$q=0.99$，$\nu=100$我们得到$L\approx 1/e$)。另一个困难涉及到复制机制。互补碱基间有识别的一定精确度要求低温。在这

些温度下所形成的双螺旋是稳定的。所以，正如一给定链的重复模板作用所要求的那样，为了分解螺旋，较大的温度梯度或者涨落是必须的。由于各个二级结构的缘故，这不会发生大的速率差异（如在现今酶指令的复制过程中所发现的）。

如果我们最后提出下述问题：

没有催化帮助，核酸能够组织起一个自我复制的并继续进化的单元吗？

我们的回答必定是：

由于它们的互补相互作用，由一条正链与一条负链构成的每一个集合具备内在的自我指令性质。在适宜的条件下，它们可能曾选择过具特定序列的单集合。不过，这些序列，即使都是可复制地出现，也只代表了非常低的信息量($\nu<100$)。由于不同的集合必须互相竞争，这种系统（没有自我指令的催化帮助）不可能将自己组织成任何类型的相关功能。

5 通过循环催化的自组织：蛋白质

5.1 酶的识别和催化

在我们探究蛋白质是否能够形成自我指令系统之前，我们应该注意到它们的一些性质：

1.正如在导言（第一章）中表明的，20种氨基酸可形成极其多样的序列(参见表2-3)，实际上偶然形成的只是其中极小的一部分。另一方面，氨基酸并没有固有的自我指令或互补指令性质。无论在什么地方，趋近于此种性质的任何东西出现时，似乎都会具有例如由(+ -)盐桥稳定化的"折叠"结构即β结构，或者具有其他诸如胶原蛋白那样由一些氨基酸互补排列的简单的规则结构，这是某种特殊排列的独特结果，但不是象核酸那样是字符的内在性质。这种特定的"指令"排列缺乏非常重要的突变诱发性。当错误出现时，它们不能复制错误拷贝。而且，多肽链强烈倾向于采取特定的空间折叠，这对任何简单的复制过程都是一种巨大的障碍。

2. 另一方面，空间折叠是蛋白质具有识别特定结构能力的基础。酶的催化性质是这种特征的另一个结果，假如识别基团

也具备协调的催化功能。作为一个例子，胰凝乳蛋白酶的活性中心如图2-10所示。只是在最近，才由x 射线结构分析精确揭示了这些功能团的空间排列。对于功能与复杂结构的相关性，它提供了一个非常好的例子，在其结构中，相距颇远的顺序排列的基团联系在一起，成为精确固定的空间排列。极其多种多样的特定识别部位，也由大量的各种抗体清楚地揭示出来了，抗体的成键能力涉及到所有抗原基因，即使半抗原基因在进化期间不与抗体接触。而且，实验室实验已表明，多肽的随机合成涉及到大量的各种常常是高度专一的催化功能（相似于随机多肽中的胰凝乳蛋白酶的功能）。这些产物并不是可复制地形成。即使某些功能确实重复出现，它们也是由相当不同的、无

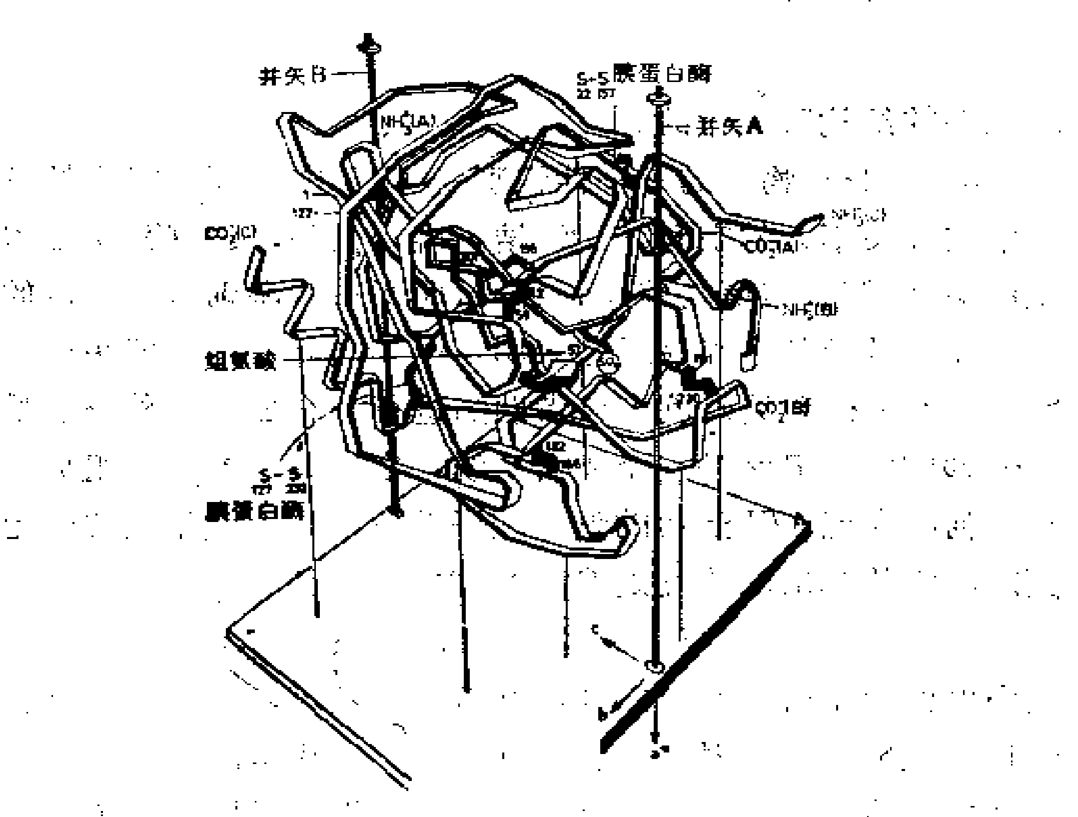


图2-10 a. α 胰凝乳蛋白酶的多肽链的构象示意图　本图取自布洛、伯克利克和喻特利的结果。

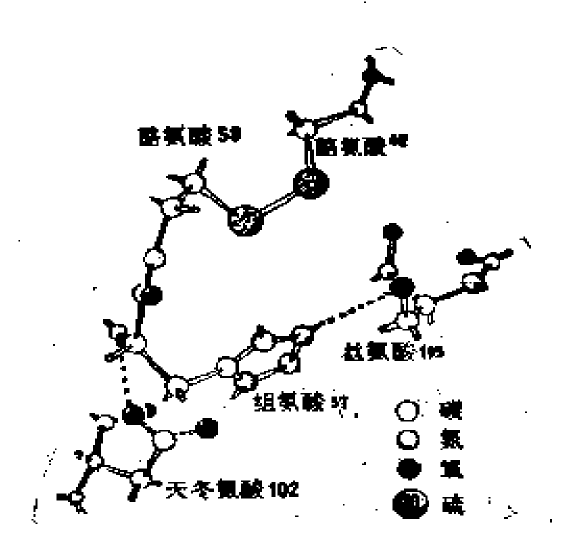


图2-10 b. α胰凝乳蛋白酶活性中心的几种氨基酸侧链的构象　显示了蛋白质的识别是一种空间折叠的独特结果，而不是字符的任何“内在”性质。〔图来源同上〕

不相关的结构所执行的。

3．大分子物种的专一性识别，被限制在范围相对小的短序列中或伸展不大的空间(三级)结构中。例如，某些酶(例如木瓜蛋白酶)降解特定氨基酸序列的肽，而另一些酶则合成特定的序列。李普曼及其小组最近已表明，一种环状的十肽即短杆菌肽S完全不需核酸编码就可以借助酶进行全合成。这种分子量为280000的酶，显然是一个复合系统，它由几个亚单元组成。无论如何，在把*ATP*激活的氨基酸联系成为如下的精确的确定序列时，它的作用的的确确就象“蛋白质模板”：

$$D\text{-}Phe\text{-}L\text{-}Pro\text{-}L\text{-}Val\text{-}L\text{-}Orn\text{-}L\text{-}Leu,$$

其中保留着蛋白质的硫醚键，直到两个这种完整的五肽环化为短杆菌肽*S*分子。因此我们必须记住，形成蛋白质时，特定的、非常精确的指令可由蛋白质单独给出，毋须借助于核酸密码。不过，这种指令限制在相对短的序列中(例如五肽)。但是，利

用这种性质可以设想一种酶的网络，它产生小肽，并经过一定的步骤使它们连接在一起，直至出现完整的蛋白质分子，同时出现的还可能有催化它们自己的再产生的功能。

4. 这种网络能够表现为高度控制的机构。由于酶功能来自于特殊基团的精确空间排列，因而与诱导物结合，或与其他蛋白质结构相互作用而引起构象变化，这种构象变化能够大大地改变酶的性质。这种关于酶功能的控制模型，最先是由F·雅各比和莫诺提出来的。昌列克斯、莫诺和怀曼，以及科什兰、尼迈塞和菲尔麦已导出了特定机制，并且已由对各种酶的动力学研究得到检验（例如，柯尔什纳运用弛豫技术对甘油醛磷酸脱氢酶进行的研究），结果表明，这些酶系统可能具有电子控制装置中为我们的全部所认识到的性质。因此，自复制网络无论何时发生进化，它都可以包括高度复杂的控制功能。

5.2 自组织的酶循环(理论)

5.2.1 催化网络

记住了第一段中勾画的性质，我们就可以构造一种“催化网络”(参见图2-11)。某些蛋白质有能力催化有限数目氨基酸聚合成特定序列的多肽链(例如，一定的五肽)；另外，这种“酶”识别这些低聚肽的一定末端序列，并使之进一步连接起来，所以最终可以出现确定的任一长度的链。参与这些催化功能的酶通常是多功能的。它们可以识别属于(各种长度的)不同肽链的特定序列，但是它们仍然是有选择性的，这取决于底物的特殊三级结构，以及识别部位的可得性，识别部位必须不被空间折叠

的肽链所掩蔽。

现在让我们假定，这个网络中的每一种催化剂都是在另一种催化剂的帮助下制造出来的。一定催化控制功能的激活，是由某种链的结合或裂解引起的。相似的过程今天已为大家熟知，例如，通过靠近链一端的肽键的裂解，从胰蛋白酶原（或胰凝乳蛋白酶原)激活为胰蛋白酶(或相应胰凝乳蛋白酶)。以这种方式，可以形成如图2-11所示的有众多分支的循环网络。由于每一种酶都需要一种以上的酶以进行自复制，故为了造成这种自

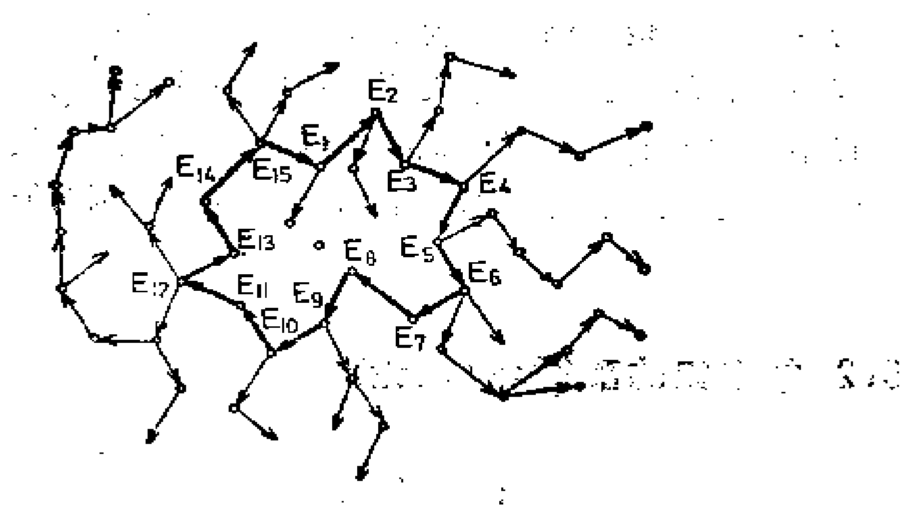


图 2-11 蛋白质的催化网络 其中包括一个封闭环：$E_1, \dots, E_{11}$.

复制网络，至少有一部分酶必须是多功能的(参见分支)。例如，如果能够识别至多为 5 种氨基酸的序列，那么产生一条80个肽键的链至少需要 5 种酶以把聚合度增至80，即：

$$E_1\ 5E_2 10E_3 20E_4 40E_5 80.$$

网络的范围越大，发现其中某个封闭环的机会就越高。只有这种环的闭合才使得系统是自催化的，并从而保护自我复制。如果环足够大，所有的辅助功能诸如产生各种各样的低聚肽和新

体链的辅助功能，就容易定位在分支上。

5.2.2 自复制环及其变体

现在，我们将集中注意那些出现在封闭环上的酶，并标记以E_1，…，E_n，就象有机化学家寻找在某种复杂芳香结构中的“生色团”。让我们用图2-12所示的循环图来代表这种环。反应速率的微分方程，一般是非线性的。不过为了简单起见，我们考虑相应于前面处理的缓冲的底物浓度的一种线性近似。纵然这不象单体字符情况中那么现实，在此网络中，循环途径上的底物主要涉及到聚合前体，但是这样的条件在原则上是可以达到的。而且，基本的结论仍适用于非线性情况。

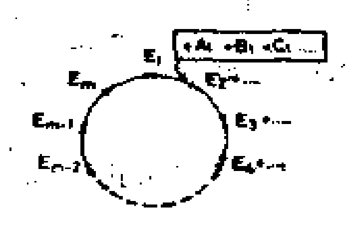


图 2-12 表示催化循环的图

速率方程组是第 4 章中给出的方程组(8)或(18)的相应的推广。对每一个有 m 个成员的循环，在没有选择约束时，我们有

$$
\left.
\begin{aligned}
\dot{x}_1 &= F_1 x_m - R_1 x_1 \\
\dot{x}_2 &= F_2 x_1 - R_2 x_2 \\
&\vdots \\
\dot{x}_m &= F_m x_{m-1} - R_m x_m
\end{aligned}
\right|
\quad
\begin{aligned}
F_i &= k_0 A_i L_i \\
R_i &= k_0 D_i
\end{aligned}
\qquad (5\text{-}1)
$$

其速率系数矩阵

$$\begin{pmatrix} -R_1 & 0 & 0 & 0 & F_1 \\ F_2 & -R_2 & 0 & 0 & 0 \\ 0 & F_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & F_m & -R_m \end{pmatrix} \tag{5-2}$$

特征方程产生关于本征值的m个根

$$\prod_{k=1}^{m}(R_k+\lambda)=\prod_{k=1}^{m}F_k \tag{5-3}$$

因为所有的R_i和F_i都为正，我们获得λ的m次多项式，多项式中所有包含$i\geqslant 1$的λ_i项具有正的系数。不过，常数项（$\sim \lambda^0$）为：

$$\prod R_k-\prod F_k \tag{5-4}$$

它对于$\prod F_k>\prod R_k$为负。在这种情况下，多项式中含有一个相反符号，即除了最后一项以外的所有项都是正的。按照笛卡儿规则，可以找到一个正的和$(m-1)$个负的（可能为复数）本征值。

类似于(4-13)式，特别假定$R_1=R_2\ldots R_m\equiv R$，我们得到正的本征值：

$$\lambda=+\sqrt[m]{\prod_{k=1}^{m}F_k}-R=\hat{F}-R \tag{5-5}$$

在不存在分解时（即所有的R项都等于零），解就简单地归结为方根，它包括了几何平均$\hat{F}$的绝对值乘以m次单位根的值，单位根依据下式：

$$e^{ik2\pi/m}=\cos(k2\pi/m)+i\sin(k2\pi/m)$$

$$k=0,\ 1,\ \ldots,\ (m-1) \tag{5-6}$$

在“平衡”后[即在伴随负的本征值的$(m-1)$个标准模衰退以后]，循环的不同成员的群体密度趋近于恒定比例。这些“平衡

比例"是（在上述R为一致的条件下）：①

$$\frac{\bar{x}_{i-1}}{\bar{x}_i}=\frac{\sqrt[m]{\prod_{k=1}^{m} F_k}}{F_i}=\frac{\hat{F}}{F_i} \tag{5-7}$$

$$\frac{\bar{x}_i}{\sum_{k=1}^{m}\bar{x}_k}=\frac{1}{1+\frac{F_{i+1}}{\hat{F}}+\frac{F_{i+1}F_{i+2}}{\hat{F}^2}+\ldots+\frac{F_{i+1}F_{i+2}\ldots F_{i+m-1}}{\hat{F}^{m-1}}}$$

$$=\frac{1}{1+\sum_{s=1}^{m-1}\prod_{k=1}^{s}(F_{i+k}/\hat{F}^s)} \tag{5-8}$$

(5-8)式中的下标是循环取值，即对于$i+k>m(k=m-i+1,m-i+2,\ldots,m-1)$，$i+k=i+k-m$。所获结果的物理学解释是：

每一个循环，具有一个代表了整体"集合"的自催化生长的简正模以及$(m-1)$个模，它们代表弛豫现象即在循环之中的"平衡"。所以，一个封闭的催化环等价于一个具有自我指令或互补指令的系统。

在我们对在选择约束下竞争性循环之间的选择行为作进一步分析之前，我们可以考虑不同的反应网络，以对选择的前提有更多的了解。

对于反应网络的选择自复制，有两个条件看起来是必要前提：

1．系统必须包含反应的"封闭环"；

---

① 对$R_1\neq R_2\neq\cdots\neq R_m$，该表达式将复杂得多，但容易用递归公式导出。从如下恒等式入手：

$$1=\frac{x_m}{x_1}\cdot\frac{x_1}{x_2}\cdots\cdot\frac{x_{m-1}}{x_m}$$

并加上"平衡条件"：

$$F_1\frac{\bar{x}_m}{x_1}-R_1=F_2\frac{\bar{x}_1}{\bar{x}_2}-R_2=\ldots\ldots F_m\frac{\bar{x}_{m-1}}{\bar{x}_m}-R_m.$$

2. 在反应状态之间的耦合必须是催化性的。

为了阐明第一个条件，我们可以考虑一条一般的开放（准线性）反应链：

$$\begin{array}{c} +\text{反应物} \\ \overrightarrow{V} \\ \Longrightarrow A_1 \to A_2 \cdots \to A_{m-1} \to A_m \Longrightarrow \\ +\text{产物} \end{array} \tag{5-9}$$

又为了简单起见，我们假定，除了物种A_i的链以外，所有继续参加反应的反应物都处于高浓度(并且因之是缓冲的)，因此对它们不必加以明确考虑。在恒反应流$\overrightarrow{V}$处，将充满不同的状态直到定常水平，整个系统终将"溢出"。如果条件适宜，系统将"产生"不同的状态，但是不能"再产生"它自已，即一定状态的产生速率与它们的群体无关。哪怕该系统有一些A_i或全部A_i是催化活性的，这即意味着它们是无消耗地参加反应的，此种系统仍然不具备重要的自催化生长性质。

现在让我们考虑类似于图2-12中的循环系统，不过，其中的E_i不是催化剂，而是简单的反应产物和反应作用物。系统可以有流的交换，除了E_i以外的所有作用物(A_i、B_i、C_i、…) 也是缓冲的。这种循环将再产生它自已。可是，只要任何一种E_i都不是催化剂，系统就将趋向一个定常状态，此时每一E_i在产生出来后又将被消耗。这样一些循环并不会互相竞争，而将采取某种平稳分布。在生物系统中的确存在这样的循环，并且这些不同的反应步骤通常也是由酶催化的。不过，这些酶不能通过此种特殊循环来复制出自身，它们的保持，是通过其他一些作为这一整个生命实体的自维持循环一部分的"回路"来实现的。

数学处理可以进一步说明这种行为。在不存在其他分解反

应的情况下，非催化的从E_i到E_{i+1}的转变过程，对于E_i的消失和E_{i+1}的出现，速率表达式完全相同。所以，在速率系数矩阵(5-2)中，所有$-R_i$都可用$-F_{i+1}$代替。

这样一来，特征方程仅仅产生出$(m-1)$个负的本征值：

[例如，对 $F_1=F_2=\ldots F_m\equiv F$；$\lambda=F(\sqrt[m]{1}-1)$；

或者对$m=4$：$\lambda_1=0$，$\lambda_2=-2F$，$\lambda_{3,4}=-F(1\pm i)$。]

注意，对于非循环反应，矩阵[参见(5-2)式]右上角的项将消失。对于上述例子，特征方程将变成$(1+\lambda/F)^m=0$，从而产生出$\lambda_{1,2,3,4}=-F$。

有选择的自复制的基本条件是，循环的反应作用物A_i不被该反应消耗，即它们是催化剂。当循环中并非所有步骤都是催化的情况下，该循环仍将保持其竞争的、自我复制的性质。换言之，不必要求反应物在每一阶段都是无消耗地催化下一步骤。实际上，整个循环所需要产生的，仅仅是能使循环变成自催化循环的催化剂。从上述处理中[参见矩阵(5-2)]，容易看出这一点。在非催化循环中，我们可能必须把所有的$-R_i$代换以F_{i+1}，因为在一个反应中，物种消失时则出现其产物。不过，如果循环中恰好是第一个物种代表了从E_m形成的催化剂E_1，E_1在E_2形成时又不被消耗，那么，矩阵(5-2)的左上方的项变为零。

在此我们得到

$$\lambda\prod_{k=1}^{m}(F_k+\lambda)=\prod_{k=1}^{m}F_k \tag{5-10}$$

它的确有一个正的实本征值。例如，对于$m=2$且$F_1=F_2=F$

$$\lambda(F+\lambda)=F^2$$

$$\lambda_1=\frac{F}{2}(\sqrt{5}-1) \tag{5-11}$$

$$\lambda_2 = -\frac{F}{2}(\sqrt{5}+1)$$

5.2.3 不同循环之间的竞争：选择

给定一定数量不相同的独立(未耦合的)催化循环，其中每一个循环都处于“内平衡”，则该系统由属于正本征值的简正模y_{i+}来标征。由于此平衡系统中所有其他的模都已衰退，所以我们可用 $y^* = \sum_{k=1}^{m_i} x_{ik}$ 代换 y_{i+}。处于恒力的每一循环的选择方程可取得以下共同形式：

$$\dot{y}_i^* = k_0(W_i - \bar{E})y_i^* \tag{5-12}$$

式中的选择价值 W_i 由此循环的正本征值给出。(例如对 $R_{i1} = R_{i2} = \ldots \equiv R_i$)：

$$W_i = \hat{F}_i - R_i \tag{5-13}$$

式中 $\hat{F}_i$ 按照(5-5)式。仅当 $\hat{F}_i > R_i$ 或者 (对于不同的 R_{ik} 时) $\hat{F}_i > \hat{R}_i$ (符号∧标征几何平均)，才能够出现选择。

与(4-23)式类似，我们还能够计算被选择的物种浓度。我们知道，只要起始反应能够用线性方程系统来描述，则独立循环的行为便相似于各个自我指令物种或互补集合。

不过，这些循环不必是互相独立的。由于分叉(例如：某些酶的聚合功能行为)，因而它们连接在一起，例如图2-13所示的例子。如果对所有物种写出速率方程，此时不考虑它们属于哪一个循环，但是明确依据循环中它们成员之间关系加以整理，那么，我们容易从速率系数矩阵导出它们的性质，如图2-3所示。

这些系统的典型性质是，若耦合起源于被选择的循环中，则被选择的物种(最大的W_i)带有所有耦合的分支和循环。一个循

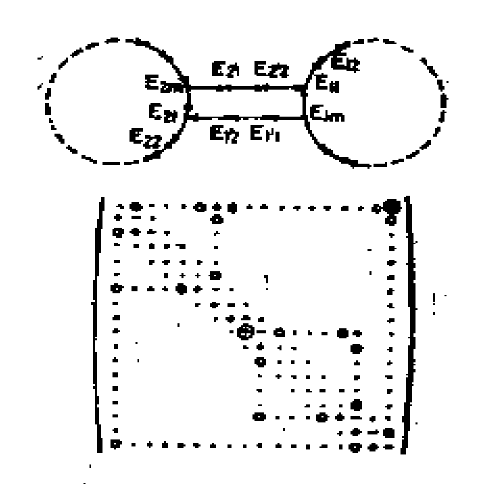


图 2-13. 由反馈回路所致耦合循环的示意图。该系统中包含了三个循环，即循环1和2再加上包括这两个循环的"大"环。由于反馈，这些循环并不为了选择而互相竞争，而是互相稳定化。总系统的速率系数矩阵反映了反应行为。其中的列代表了物种，行代表了反应。加有圆圈的两个点⊕代表了环的耦合点（即通过由E2/2和E1/2催化的反应）。如果这些元素都为零则矩阵可以分解，分解后得到的两个矩阵代表了两个独立的（竞争的）循环。框架区域代表了三个封闭环。这些框架之外的对角线代表了简单的分支。

环的分支，若它完全不具备有利于循环复制的功能，则将是非常不利的东西，因为它代表了无用的稳定因素，它限制利用复制功能，从而可能降低选择价值 W_i。我们将把这些分支称作"寄生分支"。

5.3 蛋白质能复制出自己吗？

援引前面段落的结果，便会对这个问题作出否定的回答。初看起来，对于开始自组织，蛋白质似乎比核酸要合适得多，因为蛋白质有两种性质：

1．借助于它们的三级结构在识别一些底物时所获得的明

显较高的精确度，以及

2．比起长度有限的单链信息容量，它在多步循环（及其分支）中含有更高的信息量。

不利之处在于，单独一个蛋白质不可能从单字符复制出一条长链，以及尤其在于专一性识别不是“任何”链的内在性质，而仅仅是“某些”链的独特性质，或者不同物种的特定功能性质的某种罕见的巧合。

结果造成，通过特定循环而催化它们自己复制的蛋白质，不会自动地复制它们的由错误复制形成的突变体，即使错误拷贝具有优势亦如此。由于既有有利之处也有不利之处，故我们必须对上述问题详加分析。

业已表明，独立的催化循环是“自我指令的”。这类似于诸如核酸那样自我指令或互补指令的单物种（与蛋白质相反，这是核酸的内在性质）。除了前面提到的诸如短杆菌肽S（遗传上对它编码的）那样一些抗生素的生物合成系统以外，我们目前尚不清楚自然界是否存在此种自我维持的蛋白质网络，但是我们至少能够把它们想象为人工制备物。

由它们自己形成这种循环的概率有多大？如果我们只注意催化不同氨基酸中形成肽键的功能，那么我们可以说，在一个随机的序列群体中有一些序列总将显示出这种催化活性。因此，类蛋白物质虽不具备专一性序列，但其整个生成过程是自催化的。这将是生命体进化的基本前提，重要的还要注意到，毋需核酸指令它就能出现。

不过，对于进化行为，只有非专一化自催化生长还是不充分的。只能通过利用选择优势，系统才能改进自己，而这需要序列具备专一性。第一章中已表明，找到明确指定的几个序列巧合一致的概率必定是非常小的，以致没有任何意义。另一方

面，只有最佳适应的酶（正如我们今天发现的），才能由独特的序列来代表。在远未最优化的情况下，就可以出现这种系统了，由此又说明也许只要求序列中不多的一些关键位置就够用了。有相当大的一类不同的序列，它们只有为数不多的位置是相同的，与现在的酶相比，在很大程度上，上述这些序列可能会出现某种专一性功能。

让我们假定，存在充分大量的功能蛋白质，它们对从前体到其他蛋白质的形成进行催化。其中每一种这样的蛋白质都用点阵中的一个点来代表，并在催化耦合的点之间划出连线。假定每一个都是另一个给定点的催化活性目标，则每个点具有相等的先验概率P。一个包括k个点阵点的连续环总概率为P^k。存在着$\frac{n!}{k(n-k)!}$种可能的R元封闭环。表达式$\frac{n!}{(n-k)!}$代表了由于元素的选择和序列不同（即在k类中n个元素的不重复的变体）而具有的全部配容数。由于排列是循环的（对任何诸如$DABC$和$ABCD$等的k位置，一个循环的开端和终端是任取的），所以这些变体是R重简并的。出现任何R元循环的概率是：

$$P_k=\frac{p^k n!}{k(n-k)!} \tag{5-14}$$

对大的n和$(n-k)$，我们可以应用斯特林公式

$$n!\simeq n^n e^{-n}\sqrt{2\pi n} \tag{5-15}$$

并获得P_k的极大值（依据$\partial P_k/\partial k=0$）：

$$p=\frac{1}{n-k_m}\,exp\left\{\frac{2n-3k_m}{2k_m(n-k_m)}\right\} \tag{5-16}$$

从式中可知，对任何$p>\frac{e}{n}$，有k_m接近于n；而对$p<1/n$，有k_m趋近于1。

于是得出结论，对大的群体n（对此我们可发现$p\gg 1/n$），将形成大

的环，环上有许多交连，它们几乎涉及到所有催化活性蛋白质。

上述程序假定了催化耦合有一致的先验概率，这个假定也许是成问题的，因为它过度简化了实际上是相当复杂的情况。若要更精确地描述实际情况，则需要比目前已知的更为详尽的信息。也许在一定的序列和一定的复制功能之间正好有特殊的关联，但是它们仍不会象核酸那样是结构所固有的性质。在核酸中一定的序列总是要诱发它自身的复制。蛋白质缺乏这种序列和复制功能的先验关联，所以关于复制环的结论只是在定性上有效。当然，每一耦合都由特定的速率参数来标征，这就如选择价值W_i所表达的那样。出现特定耦合的概率将随W_i值增加而减小。因此，我们实际上应该运用概率分布函数$p(W)$，并说明在W和$W+dW$之间发现某个具有选择价值的耦合概率$p(W)\cdot dW$。如果由不同选择价值W_i标征的不同环为选择而进行着竞争，那么这种说明将是必要的。每一个高W值的环都将与许多具有较低选择价值的环(伸展了的)相联接，而不可能通过选择而"逃脱"这种联接。

简单复制环的一个主要缺点在于，它们不可能避免选择低效的寄生耦合。

对于进化，催化循环还有另一个甚至更为严重的不利之处——假定了存在某个与竞争者对抗而被选择的独立循环：

这个循环(图2-14)，可能由通过该循环中包含的酶的不精确复制所产生的突变体而进化。设E_1'是一个这种突变体。难以说E_1'正好是比E_1"更好的"催化剂，因为它将再产生$E_2, \ldots, E_n$，并从而再一次导致E_1而不是E_1'。为了改进，该循环必须经历一条特定的突变体链$E' \to E_1' \ldots E_n' \to E_1'$，即它必须形成一个新的特定循环。出现一条特定事件链的概率，比之出现"任何"k元循环的概率，的确要低得多(在下图中，按照(4-14)式为p^k比P_k)。

尽管在一般条件下(反应物未缓冲的条件下)催化网络将是

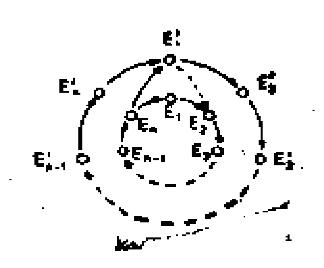
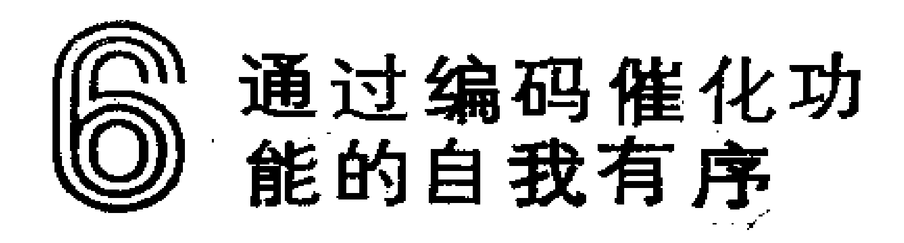


图 2-14. 在催化循环中突变体的复制

非线性的，但是我们至此为止已用线性速率方程描述了这种系统。而且，速率方程(5-1)所描述的行为上，也许已加上了特定的非线性控制效应。于是，引起的问题是，所考虑的非线性效应是否会彻底改变那些基本结论。我们肯定将观察到解的性质的彻底改变(作为例子，可参见下面的段落)。然而，由于缺乏内在的自我指令或互补指令，所以系统不可能进化这个基本结论，对于非线性催化网络更是适用。非线性催化网络中的可复制性又是独特的大分子序列的巧合的结果。此种系统，由于负担了过多的“低选择价值信息”，而不能方便地利用“选择优势”。选择行为的“线性”近似在这个方面是相当有代表性的。



6 通过编码催化功能的自我有序

6.1 核酸和蛋白质需要协同作用

我们把核酸和蛋白质作为信息处理分子的原型进行研究，结果得出如下一些结论：

1．核酸对自组织提供了一种重要前提，即互补指令，它是作为“固有的”选择自复制，以及运用偶数(例如二元或四元)字符系统的密码信息的基础。识别能力不可能达到这样高：允许在单链中积累起大量的且是可复制的信息量。这将要求此链中的单互补字符有相对较高的相互作用自由能，而反过来这会使得密码过于“粘性”，难以用合理的速率进行阅读。幸亏字符相互作用的协同性，密码子—反密码子识别才可以是相当有选择性的，并仍能在几微秒内加以完成。由于单链中的互补相互作用，所以能够形成特殊的各个单链结构(作为酶识别的目标)。即使始终存在某些催化能力，也是太弱了，难以引起竞争信息载体之间的耦合，从而不可能导致信息量的显著增加。

2．蛋白质，与核酸不一样，它正好具有这种性质，即具有巨大的功能的和识别的多样性和专一性（例如可参见抗体的识

别能力）。通过催化耦合，它可以把许多信息载体连接在一起，从而构造出非常之大的信息容量。不过，识别不是其中序列元素（例如，氨基酸残基）的固有性质，而是活性部位的空间结构中残基相互作用的某种特定巧合。这种结构可以可逆地改变、修改识别（以及催化）能力，并由此提供了控制性质（例如通过"变构"激活）。蛋白质亦可显示出一般的催化行为，以及通过循环耦合而选择。可是，这并不是所有蛋白质分子的某种"固有的"性质，而是某种特定物种的独特性质。所以，蛋白质不容易利用在"表现型的"突变体中出现的选择优势，并因此缺乏进化所要求的、而核酸能够提供的基本前提。

3. "线性"反应系统不可能把为自组织系统集结所必要的所有性质都结合起来。如果信息载体是竞争性自复制的，那么选定的信息容量就被限于一个物种的信息容量。它们需要催化耦合，以扩大它们的信息容量。另一方面，如果复制仪是由（线性的）循环催化耦合带来的，那么它可以包含大量信息量，但是该系统不可能避开"寄生"分支而选择。只有非线性系统（进一步的论据参见后面），才能提供开始自组织所需的全部性质，并允许系统继续向高水平进化，直至能够逃脱它起源时的特定前提。（这种高级水平不可能由任何有限概率的随机装配来达到。）

4. 互补指令与催化耦合的结合将导致非线性选择行为。我们必须找到耦合核酸和蛋白质功能的最简单途径，以便重视某类能够导致活细胞结构和功能的进化行为。我们不应该妄求解释进化的历史途径。我们应尽力做到阐明最起码的前提，从而对进化过程的物理学原理有一些认识，与其特定结构无关。这种系统必须利用核酸形成密码的性质，以及蛋白质的催化能力。所以，这就需要集结成翻译机构。因此，我们必须处理两个并非互相独立的问题。

a) 某种密码及其相应的翻译机构何以能够产生? 现在,这种机构中包含了采取tRNA分子形式的连接物,以及由氨基酰合成酶所代表的识别酶。

b) 由核酸和蛋白质系统所代表的这种系统,如何能将它自己组织成稳定自复制的、继续进化的单元?

让我们从问题b)入手,因为这个问题的回答包括在问题a)的答案中。

6.2 自复制的超循环

6.2.1 模型

我们考虑图2–15中描绘的简单模型。它由一些链长有限的核苷酸序列I_i(或最好说是(±)集合)构成。与由一条或两条有催化活性的多肽链(图中以E_i表示)相比较,这些序列不必提供更多的信息。环绕I_i的环,是核苷酸集合的互补指令能力的示意图,核苷酸的每一个集合都是互复制的"正"和"负"链构成的。前面一个的多肽链E_{i-1}提供的特定催化的放大有利于这种相互复制。这个多肽E_{i-1}是由核苷酸链E_{i-1}编码的。翻译系统的存在充分保证了从I_i到E_i的精确翻译。贮存在每一I_i中的信息,只有一部分必须用于为增强放大功能而编码,以有利于下一个信息载体的生成;其余部分保留用来编码诸如翻译聚合、控制功能等等一般的酶功能。而且每一具备对下一个信息载体的生成有特定的放大功能的E_i,都可以(但不是必须)是一个特定的聚合酶(例如,$Q\beta$聚合酶,参见第7章)。它也许正好是作用在某种普通的聚合酶上的某个特殊诱导子(或抑制子)。重要的是,

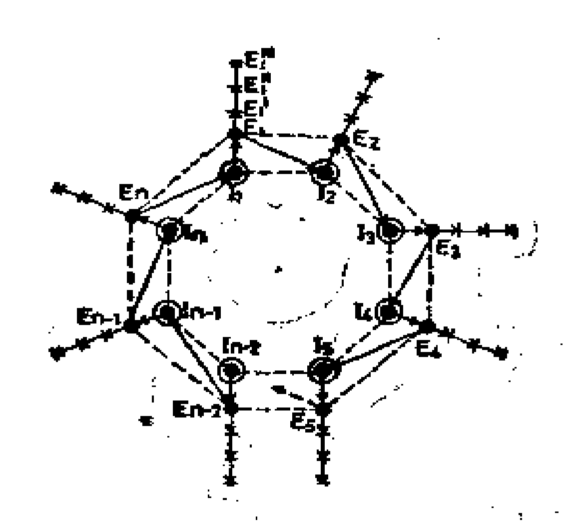


图 2-15 自我指令的催化超循环。

I_i 代表信息载体，即互补的 RNA 单链。每一个小循环都表明包含两条互补链的 I_i 集合的自指令性质。E_i（由I_i编码）代表催化功能。每一 E_i 分支可包括几种功能（例如聚合、翻译、控制）。其中一种功能必须提供与信息载体 I_{i+1} 的耦合（例如，由特定识别而促进 I_{i+1} 的生成）。表示所有耦合的连线必须是封闭的，即必定存在促进 I_1 生成的E_n。此种超循环由非线性微分方程系统描述。

整个"超循环"是封闭的，即有一个对I_1反馈的 E_n。从而，此种系统代表了一种"循环等级组织"，其中许多循环的(互补)核苷酸集合由酶的"超循环"联接在一起。这个二级封闭环是重要的，因为若非如此，则不同的I_k不会协同而将竞争，从而互相对抗地进行选择。

这种系统具备下述性质，这些性质在后面还要详加讨论：

1．每一个循环，就象第 4 和第 5 章中的系统，都具备自催化生长性质。

2．独立的循环为选择而竞争。

3．作为非线性的结果，选择将是非常明确的，如果包含有奇点，则很可能象"全有或全无"行为。

4．由于具备这些选择性质，此系统将能够：

a) 利用非常小的选择优势(在随机意义的范围内它必定出现),以及

b) 非常迅速地进化。一个被选择的系统将不允许独立竞争者集结，所以密码和手性将是普适的。

5. 循环耦合将提供适合于系统要求的信息容量。然而,单个密码单元的复制长度(参见表2-8的ν_{max})将充分小以保证复制性。

6. 此系统能够进化,即能够由利用选择优势而改进。"基因型的"突变，即在I_i中的变化立刻能够被E_{i-1}利用。为了传遍整个循环，因而不必等待有关的突变系列，而这在"线性的"催化循环中是必要的(参见第5章)。选择优势，通过抑制、解抑制或启动，可以变得有效。

7. 如果寄生分支的选择价值小于循环成员的选择价值,那么系统将对抗寄生分支而选择。如果具有较高选择价值的寄生分支从一开始就存在，那么它们将不允许该循环集结。但如果它们在此循环集结以后才出现,作为非线性选择行为的结果,它们便没有生长的机会。一个循环通过收缩可从多个成员减少成一个成员，如果这提供了某种选择优势的话。对于耦合系统的同时存在，所需要的条件是相当有限的。

8. 只有一种类型的分支可以与循环共存,此种分支的选择价值正好与循环的选择价值相匹配。如果只有在循环内的分支和编码区域I_i利用位于E_{i-1}上的同一个启动子,那么才可能正好相匹配。这将自动地导致密码系统的基因和操纵子结构。在这些分支之中，系统能逐渐形成一般的应用功能(例如，聚合酶、翻译系统、控制因子、代谢酶)。

9. 这种系统在集结之后，立刻溜入某个分隔间。只有分隔间化的系统才能排他地利用功能分支(由突变带来的)，并使之

成为自己的优势(并且从而允许分支进化)。同样的机制保护了系统免受不利分支突变带来的污染。

10．关闭在某个分隔间中的系统，可以由把其密码单元联接到一条稳定链上而“个体化”，即借助于某个(进化的)连接酶而“个体化”，并将整条链作为单独一个单元而复制出来。在这种链中(如果包含连接酶，那么它将是循环的)，从一定单元 I_i 形成的基因应该定位在相邻位置中。不过，在 I_{i-1} 中出现的耦合因子的讯息，可能被置于相当远的位置。

上述的后两点，肯定不是此循环的固有性质。它们表明，此循环并非进一步进化的“末日”。任何将使它进入类似于活细胞那样的高级水平的优势，它都能加以利用。或者换言之，只有那些设法达到分隔间化和个体化的系统，最后才有生存的机会。

6.2.2 理论处理

关于上述循环系统的理论，我已经与舒斯特共同完成了。借助于计算机程序，已进行了数据评价。这个工作将在其他地方全部发表。在此，只概括几个初步结果。

让我们首先考虑一种简单的、有局限的非线性速率方程，它对所期望的解的类型提供了有启发性的理解。若蛋白质 E_i 与其密码单元 I_i 处于准平衡，我们则不必直接考察它们的生成速率，并可以得到与其浓度比值为 y_k 即指令性密码单元 I_k(这里简单地当作自我指令)的表观浓度。而且，如果两种浓度 (E_{i-1} 和 I_i) 都不出现大量过剩，那么则可以假定在 E_{i-1} 和 I_i 之间的复合体的浓度与两种浓度的产物(相应于米凯利斯-门顿机制的二次项范围)成比例：在没有选择约束时，密码单元的生成速率方程可以写成：

$$\dot{y}_i=(F_i+F_i'y_{i-1}-R_i)y_i \qquad (6-1)$$

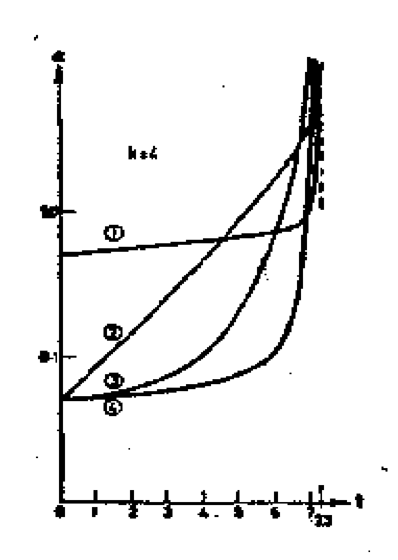


图 2-16　4元超循环的生长

（4 个信息载体 I_i 编码 4 个耦合酶 E_i）

假定在 $t=0$，$I_1$10 倍过量于 I_2、I_3 及 I_4。所有 4 个成员的生成速率由简单的二次项 $x_i x_{i-1}$ 来表示。分解则略而不计。图中，时间轴已压缩了，在此速率常数 $k=1$（或 $t=k_i'$）。在 $t=7.3$ 处出现奇点，如果包括几个竞争循环，则导致非常确定的选择。（计算是由舒斯特进行的）

（所有的高能单体字符再一次假定得到缓冲。"逆流"项也略而不计。）

这个方程的解类似于(2-21)式，在有限 t 处具有一个奇点（假若 $F_i + F_i' y_{i-1} > R_i$）。图 2-16 示意了一个例子。有趣的是注意到，对 3 元或更多元循环，可以出现振荡。有选择约束时，这种振荡甚至更为显著（参见图2-17）。2 元循环不出现任何振荡；3 元循环则出现振荡，但是仍有衰减的振荡；对于 4 元及其 4 元以上的循环振荡，衰减消失了。如果我们从与时间无关的平均出发（相应于内部平衡），那么振荡由小扰动造成。从平稳解（内部"平衡"可获得与时间无关的平均值：

$$\frac{\langle y_i \rangle}{\sum_k \langle y_k \rangle} = \frac{1 - \sum_k \frac{b_{i+1} - b_k}{b_k} / \sum_k \langle y_k \rangle}{a_{i+1} \sum_k \frac{1}{a_k}} \tag{6-2}$$

式中求和遍取循环的 k 个成员，按(6-1) 式 $a_i = F_i$ 且 $b_i = F_i - R_i$。符号

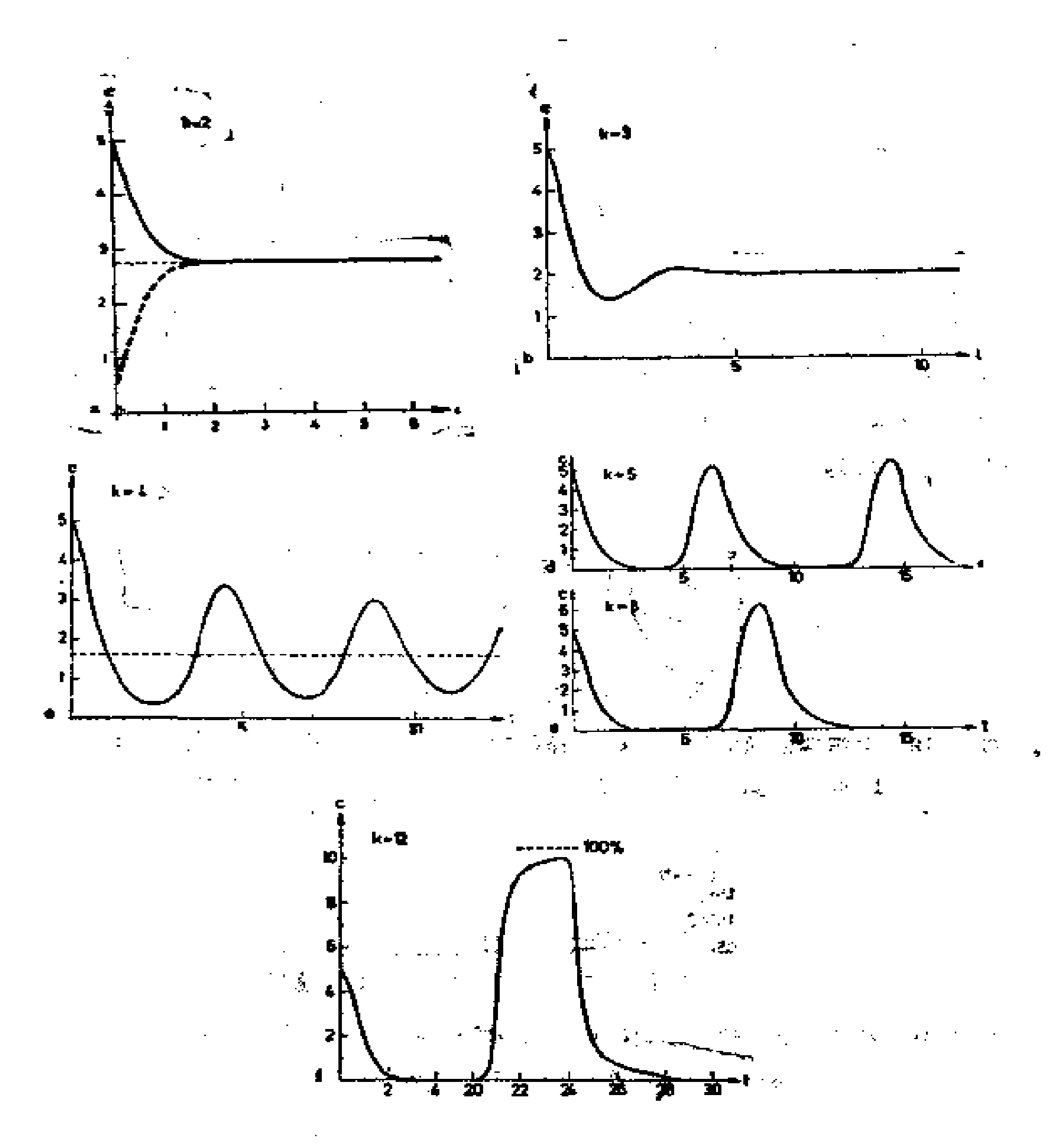


图 2-17 这些解描述了在恒总组织约束下的k元超循环选择（$k=2,3,4,5,6,12$）。此反应系统还由对所有成员都一致的简单二次生成项（参见图 2-16），以及为保留条件 $\sum_k x_k=$常量的一次"消失"项所描述。这些解只表示了其中一个成员。对 $k\leqslant3$，"平衡"值是恒量。对 $k=3$，由衰减振荡代表趋近选择平衡；而对于 $k>3$，则出现平稳振荡。要表明这种情况，可以从恒定分布入手，并在 $t=0$ 时引入某个微扰，从而造成振荡。（由舒斯特计算）

〈 〉表示振荡浓度的时间平均。

随着密码单元数 k 的增加，振荡行为以具有特征性形状的尖峰波形出现，这种尖峰贯穿整个周期。这些波峰的形状依赖于 k，还依赖于各个密码单元的生成和分解的速率(参见图2-17和图 2-20)。竞争着的循环的选择行为，不仅依赖于该循环的平均速率参数，而且依赖于各个单步骤及其不同步骤分布的相对速率，还依赖于循环中密码单元的数目 k (图2-21至图2-23)。对此种竞争的结果进行猜测将是困难的，也几乎象猜测赛马赌注一样扣人心弦。在本文范围内，不讨论此类反应机制的细节(读者可参阅其他书籍)。对耦合的循环行为亦是如此，其性质在许多方面类似于社会行为。

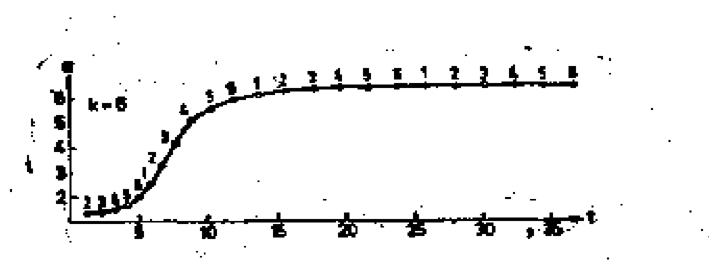


图 2-18 在振荡的超循环(k=6)中振幅的迁移 这个超循环是从在 $t=0$; $x_1^0=1.35$, $x^0_{k>1}=1.25$ 处的一个恒定的、平稳的分布建立起来的。

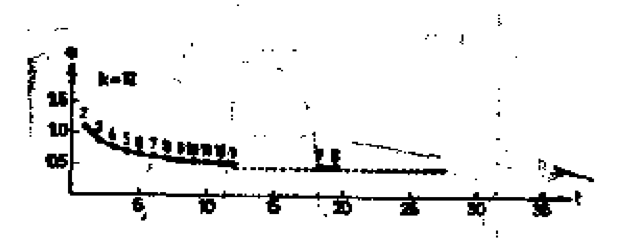


图 2-19 作为比较：线性催化循环中的振幅衰减 参见第 5 章的讨论。在选择约束下，这些浓度总是衰落到某个恒定的定常水平。

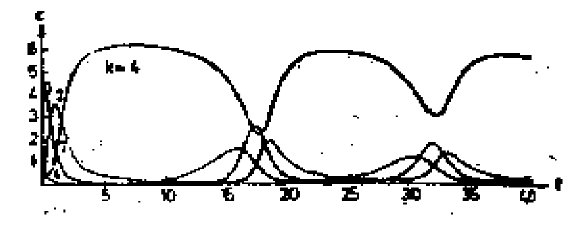


图 2-20 具有不对称速率分布的平衡的 4 元超循环中的平稳振荡 第 4 个成员的生成速率比其他成员的生成速率小 10 倍。

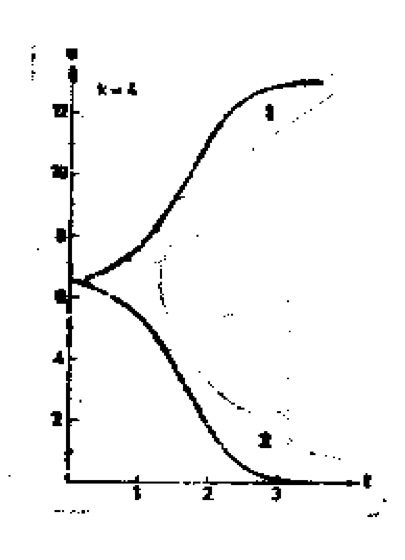


图 2-21 两个 4 元超循环的竞争 其中一个处于不利地位，这反映在生成速率常数上有 10%的不同，即：

$F_{11}'=F_{13}'=F_{14}'=1.0$；$F_{21}'=F_{22}'=F_{23}'=F_{24}'=0.9$

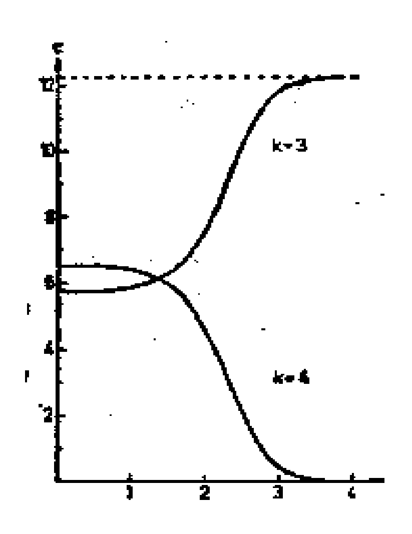


图 2-22 3 元循环与 4 元循环的竞争 两个循环的各个速率参数相同，但起始浓度不同。结果是3元循环"获胜"。3元循环：$\sum_k x^0_k=5.76$；4 元循环：$\sum_k x^0_k=6.50$。注意，4 元循环中的每一物种比较 3 元循环中的每一物种，起始时浓度约低 15%。

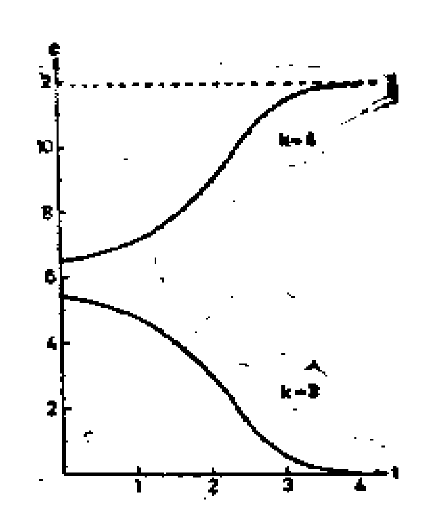


图 2-23　3元循环与 4 元循环的竞争　两个循环的各个速率参数相同，但起始浓度不同。3 元循环：$\sum_k x^0_k = 5.4$；4元循环：$\sum_k x_k^0 = 6.5$。注意，这里是 4 元循环获胜，尽管 4 元循环的每一物种比较 3 元循环的每一物种，其起始浓度仍旧是低一些。

更完全的解必须考虑如下复杂情况：

1．系统 E_i 尽管已与 I_i 耦合，但仍然具有它自己的特征生长速率。这样，我们就需要两套浓度变量，即对于酶 E_i 的浓度变量x_i和对于密码系统 I_i 的 y_i。

2．密码载体的复制速率并不与产物 $x_{i-1}y_i$ 成简单比例。米凯利斯——门顿近似（假定其中底物对酶大大过剩），在此是不能满足的。由 z_i 代表的E_{i-1}和 I_i 的复合物浓度必须运用质量作用定律来计算。这可表示为如下形式：

$$z_i = \sqrt{x_{i-1}y_i}\, tg\frac{\alpha}{2} = \frac{x_{i-1}+y_i+K_i^{-1}}{2}(1-\cos\alpha) \qquad (6\text{-}3)$$

其中

$$\sin\alpha = \frac{2\sqrt{x_{i-1}y_i}}{x_{i-1}+y_i+K_i^{-1}} \qquad (6\text{-}4)$$

上式中，x_{i-1} 和 y_i 分别指的是 E_{i-1} 和 I_i（不论它们是自由的还是复合的）的总浓度，K_i 指的是 E_{i-1} 和 I_i 生成复合体的稳定常数。

3．I_i 的生成速率包括两项，一个是线性项，另一个是与 z_i 成比例的项。总有一些 I_i 是可复制的生成，它不需要 E_{i-1} 的特定帮助。如果其中包含二次项，则循环的集结将是一个很不可几的过程。

4． 耦合的速率方程系统可以表示为：①

$$\dot{y}_i=(F'_{I_i}-D_{I_i})y_i+F''_{I_i}\cdot z_i \tag{6-5}$$

$$\dot{x}_i=F_{E_i}\cdot y_i-D_{E_i}\cdot x_i$$

其中 z_i 的意义是根据(6-3)式。

（在恒总组织约束下所获的）这个解显示了和同上面讨论的二次项范围有相当类似的性质，它对以下情况有效：

$$x_{i-1}+y_i\ll K_i^{-1}.$$

如果其中的一项浓度超过了 K_i^{-1}，那么解变成准线性的，振荡消失；但是，这一般发生在系统经历非线性区域时已经发生了确定选择的范围中。

关于竞争的循环选择行为的理论结论可以概括如下：

在选择约束下，不同的超循环将为选择而竞争。只有一个系统能生存；这个系统由具有最大价值的函数来标征，此函数具有相当复杂的形式，它能够用速率参数、品质参数以及各成员的平均浓度来表达。只要系统正在通过非线性区域，单个成员的浓度就可能振荡。选择是非常明确的，从而解释了密码和手性的唯一性。循环为了能复制出它的功能特征，必须选择一定的密码和翻译机构（参见下面）；无论何时出现这样的循环，因为在稳定的循环已经进化后，新的循环不可能与它共存，所以特定的选择行为将导致对这种特定密码的普遍利用。对于手性亦是如此。一旦聚合功能已碰巧偏向于某种立体专一构型，它们就将继续下去，使之进化得完善，这种完善要求一致的立体专一性。虽然两种构型都有“先验的”相同机会，但是由于涨落，其

① 区别自由的浓度[即分别是(y_i-z_i)或(x_i-z_i)]，并区别束缚物种(z_i)，这并不会改变方程的一般形式，因为这些项与 y_i(或 x_i)及 z_i 两者成比例。

中一个碰巧在集结过程出现，于是作为非线性放大的结果，它总是受到偏爱。对于具备“全有或全无”选择行为的非线性系统，只有一类能够获胜。这不一定适用于那些在非线性耦合中没涉及到的功能。例如，不是由具备编码功能的生物合成机构所制备的低聚肽也能利用 d 氨基酸(参见第 5 章提到的短杆菌肽 s)。

非线性也是对抗寄生分支进行选择的原因，在稳定的超循环已经集结后，寄生分支是不可能生长的，如果它们不是由该循环编码的某个单元的一部分循环的振荡行为，特别是对于选择的明确性方面，将增加新的特征。各个成员的复制，以循环出现的波的形式出现；而且正如在图 2-21 中见到的，在若干个循环波周期之后得以完成选择。这种行为使得耦合的超循环难以共存——如果不出现非常特殊类型的耦合。

超循环的起源有赖于翻译机构的存在，它包括一个集结步骤，这一步骤是以与5.3中的线性循环[参见(5-14)式至(5-16)式]相同的统计学类型为基础。不过，与循环的蛋白质网络不同，超循环能够进化，从而适应最佳功能。首先，它不象第 5 章中讨论的线性网络那样产生分支。其次，只要出现在循环内的选择优势代表了关于基因型突变体的复制优势，它就能够利用这些选择优势。这种进化还可以包括把此循环限制在最佳尺寸。这个尺寸必须提供充分大的信息容量，其中包括诸如聚合、翻译和控制(稍后还有更复杂的代谢功能)等所有的辅助功能。在把不同的单元 I_k 联接进入某个信息容量相对大的集合中，它有一定的优势。此种单个单元必须包括不多于两种的功能：识别功能，以便提供 E_k 系统的循环耦合；以及(偶尔地)提供某种(翻译或聚合的)辅助功能。如果循环中存在的所有功能必须从一种连续的核酸序列出现，那么从一开始就需要非常高的识别精确度(即某种已高度适应的酶功能)，否则整个信息量不可能以稳

定形式复制；它将漂移而去，直到它的所有有用信息丧失殆尽。

另一方面，循环一旦发展出充分精确的识别系统，出现把不同单元联接成为一条可复制链的连接酶，就将提供一定的选择优势。此时，相对于单链RNA，DNA结构的进化也将提供优势。超循环借以成为真正“自复制”系统的“个体化”，它必然联系着“空间隔离化”而出现。“个体化”和“空间隔离化”都不是超循环的固有性质，例如，如同上述的其他性质。不过，它们在集结后出现时，它们可能提供某种选择优势，从而是该超循环的不可避免的进化结果。其优势在于利用突变体。如果某突变体——特别在辅助功能方面——具备优势，它就将被整个“分散的”系统所利用，从而并不特别有利于该突变体的复制——如果它不是出现在分隔间之内。但如果它出现在分隔间内，就有利于已经个体化了的特定系统，使之也将对抗其前体而选择。这对于不利的突变亦是如此，如果这些突变体出现在分隔间中，它们只会影响(或甚至摧毁)它们自己的特定分隔间，从而消失，否则它们将污染整个系统。

适当的分隔间可以是如同奥巴林描述的凝聚层，或是象福克斯及其他人所揭示的原始条件下氨基酸随机凝聚所自发形成的微球体，这些微球体由类脂即主要带有疏水侧链的类蛋白所构成。类脂微球体的出现是一件很普通的事，许多作者都已指明了这一点。它们能够再产生的形成，并可在实验室实验中加以利用。

6.3 关于密码起源

正如 6.1 和 6.2 节所描述的，超循环存在的可能性显然取决

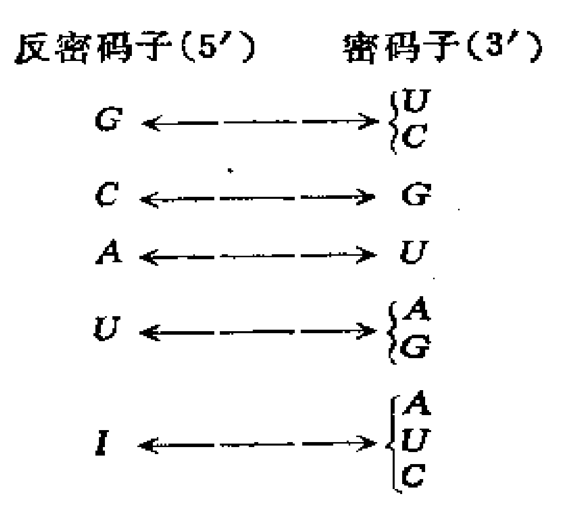
于密码和翻译机构的存在。关于与核酸相联系的内在指令性能与蛋白质3维结构的潜在功能之间的相关性的任何模型，都需要这种密码。因此与任何特殊的自我维持的催化超循环模式的存在问题相比某个独特密码的（它的存在是一个事实）起源问题，便具有更普遍的性质。

现在的遗传密码表(参见第1章的表2-2),对密码起源提供了某种线索吗？许多作者，特别是沃西的专著详细讨论了这个问题。可以提出如下的线索：

1．今日所知道的所有事实都表明,此种密码是通用的。

2．现在的三联体密码不象是从利用二联体或单联体的前体密码进化来的。正如克里克已清楚地指出的，密码子大小的变化将导致已收集起来的信息完全丧失—如果信息中的密码子所占空间不守恒，则并不提供从旧的密码子序列到新的密码子序列的简单翻译，这并不排除在密码子中不同位置的一个个单字符在它们的信息量方面有不同的影响。

3．克里克的摆动假说,对密码子中第3位置(3′位)的明显的密码简并性提供了解释。它涉及到反密码子5′位和密码子3′位的碱基互补性中的一些简并性：



这导致了在密码的第3位上的U和C以及A和G的等价性。

4．在密码子三联体中的中间位置，对于决定氨基酸的本性(疏水性、极性或电荷种类)似乎是主要的。对于以某种特定的核苷酸——氨基酸相互作用为基础的密码起源，已公认这种规律性可能有意义。

5．密码似乎反映了某个最优化原理，这也得到了公认，赖欣伯还对此进行了形式处理。对于多数密码子，3个字符中只要有1个发生了变化，就会导致氨基酸的本性发生一个小的变化，由此亦反映出关于中间位置的规律性。4类氨基酸又被分为：疏水的、极性的、荷正电和荷负电的。还可以再加上功能关联，诸如氨基酸侧链的某种酸或碱功能，以及苯丙氨酸—酪氨酸相似的结构相关性等等。对于密码的随机出现，任何最优化原理都将具有特殊的意义(参见后面)，并将影响集结过程中的不同选择。对于密码的部分重复的阅读(酶处于尚未适应的阶段也许会相当频繁地发生这种情况)，它也将对(依据各类氨基酸的)碱基序列的信息量提供某种不变性。

6．由A和U字符构成的8种密码子，比起8种GC三联体(参见第1章的表2-2)，提供了丰富得多的功能。李普曼最先强调了这一事实以及它与某种简单的密码起源的可能关系。支持这种观点的根据是：

a) 在原始条件下，A的丰度可能比之U、C和G的丰度要高得多，这导致了AU配对的丰度比GC配对要高。

b) GC配对比之AU配对具有较高的稳定性，这就使得一旦它具有选择优势时，允许GC配对在后来取代AU配对。

c) 不久前从线粒体和叶绿体中发现核糖体RNA中AU含量相对高；按照这两种细胞器代表真核细胞中的前核内含体的假说，这两种细胞器也许未曾暴露在高选择压力中。

d) 最后，把统计观点当作实际论据来看，任何密码的集结，当字符的类数尽可能小时，它是最容易的。

如何能够使氨基酸与密码子或反密码子之间发生一种确定的分配关系呢？

毫无疑问，最简单的解释将认为，这在于两套字符集之间的某种特定的相互作用。已经提出了许多模型，其范围从单密码子—氨基酸相互作用，通过氨基酸插入相应的密码子—反密码子序列互补(双)链之间的缝隙，到最后氨基酸被大的连接物分子的空间折叠结构所识别，大的连接物分子即tRNA前体或者倾向于稳定性的这种氨基酰tRNA前体。① 任何此类模型是否有价值，完全取决于是否能提供适当的实验证据，但是迄今实验证据还十分稀少。

明显的是，如果相互作用不局限在单字符之间，即不局限在氨基酸和密码子或反密码子的区域，那么能够大大增进氨基酸—核酸识别的专一性。tRNA为什么是比较大的分子，原因之一在于，它或者提供充分的特征三级结构供某些酶去识别，或者为了象酶那样识别而利用它的扩展了的结构。这两种情况都可以联系着"随机"模型来讨论。由于反密码子区域的所有排列都能够出现在其他的未变化的连接物结构中，所以困难仍在于如何防止随机的分配。

去发展任何一种只依赖于某种假设的、没有实验证据支持的模型，都是没有意义的。对于某种独特的开端，无论如何是需要相当有特色的相互作用(q接近于1)；值得怀疑的是，能否始终一般地存在充分强烈的内在相互作用，这种相互作用是直接的反密码子氨基酸分配所要求的②。因此我们有理由提问：

在缺乏特殊的相互作用时，某种独特的密码分配也能够始于氨基酸与反密码子的某种随机结合吗？

密码子(或者任何内在地与密码子相联接的结构)与氨基酸之间，一旦存在任何特定的相互作用，就会增加出现其他非直接翻译的概率。而且，我们所感兴趣的，主要不在于某种特殊的(推测性的)模型，这种模型是用于估计现在已知的连接物识别系统的可能前体的随机开始(或集结)之概率。换言之，我们希望知道，假定没有任何内在的相互作用能偏向于这种选择，那么这种复杂的系统如何能以有限的概率集结并开始独特的翻译。

让我们仅仅考虑那些已能够提供证据的相互作用。例如，我们知道，tRNA 或类似的结构能够被蛋白质 3 维结构相当专一的识别。我们还知道，氨基酸可以被(例如由 ATP)激活，并连接到核苷酸序列上；但是我们知道，并没有任何特定的、固有的氨基酸一反密码子分配能够在没有酶的帮助下满意地起作用。

关于"随机"开始翻译的最简单模型，本质上以氨基酸的密码子或反密码子分配具有相等先验概率为基础。这样一来，任何一种氨基酸 a、b、c、…可以先验地耦合任何一种连接物 A、B、C：…，a-A、b-B 等等分配仅仅是最后的结果（为此，回顾一下已经采取了的命名法)。应该提到 3 种模型，这种假定适用于它们，因此它们也从属于完全相同的一般性处理。

1. 氨基酸被类似于 tRNA 前体的聚核苷酸的三级结构(例如某条缝)所识别。反密码子定位于某个暴露环上；但是连接物上的氨基酸的识别和固定中可以不包括它，所以猜测任何三联体都可能已与给定的氨基酸发生了耦合。

---

① 与奥格尔的私人通信。

② 还有进一步的观点：如果在氨基酸和密码子之间存在这样的内在相互作用，就可以假定酶在利用这种相互作用上应该是已取得了进化，现在已该允许逆向阅读了，即从蛋白质到核酸的阅读（这与分子生物学的"中心法则"相反)。尽管有了[illegible]逆向转录(即 RNA→DNA)，但没有证据显示出(这也是很难想象的)这种逆向阅读能出现在翻译中，不过，如果它存在的话，它就该具有优势。

2．除了氨基酸用低聚肽或多肽所取代（第2连接物）而且末端的氨基酸要被激活以外，这个模型与模型1是相同的。比起聚核苷酸与单氨基酸的相互作用，这种聚核苷酸与多肽的相互作用可以更为密切、更有特色，但是对于一定的相互作用的多肽和聚核苷酸结构，任何末端氨基酸和暴露的反密码子又能够被取代。

3．这个模型类似于模型2，不过，其中的多肽具有酶功能（作为氨基酰合成酶前体），它能够专一识别将与给定连接物结构联接起来的一个自由氨基酸（或低聚肽），并且又允许反密码子氨基酸的任何可能的结合。

在上述的任何一个例子中，我们有λ种字符，即λ种氨基酸a、b、c、…，它们必须借助于连接物A′、B′、C′…（与A、B、C…互补）翻译成为密码子A、B、C…。共有λ^2种规定，例如，对于$\lambda=z$有：aA′，aB′，bA′和bB′。各种不同的分配所以可能，这在于，例如，A′是一类连接物，它们的共通之处只是有反密码子A，但是，它们另一方面却能与不同的氨基酸或激活酶发生相当不同的相互作用。假定任一分配都具有相等的先验概率。如果我们现在考虑某体积元①，在其中有在λ种这种分配，找到某个给定集合的概率（P）与所有可能集合数的倒数成比例，所有可能的集合数是λ组的λ^2个元素的“反复变更”数：$Vr=(\lambda^2+\lambda-1)$这也允许其中所有的分配是相同的那些λ组合，譬如a-A′，还允许那些其中的某个给定的连接物连接在许多不同的氨基酸上的（a-A′、b-A′、c-A′…）、或者反之亦然；总之，任何一群分配都是允许的。这是一种相当极端的（并且可能是相当不现实的）假定，但是利用它以获得较低的概率极限，所以任何偏差都只会加强这种论据。

在所有的分配中，有$\lambda!$种独特的分配，对于它们，一个给定的氨基酸仅仅联接在一个（反）密码子上，并且反之亦然。因此找到一个具有任

① 调整体积元的大小，使得在其中可找到λ种规定。这个体积元中当然有大量多肽和聚核苷酸序列。不过它们不具备把氨基酸固定在适配器上的功能。有一个问题：给定序列的规定必须是特定的；或最好这样说：该特定的规定必须是决定性的。另一方面，一个给定的结合将不可能是唯一的。大的“噪声”水平将强加于其上。

一独特分配的体积元的概率是

$$P_{ASS}\sim\frac{\lambda!}{Vr}=\frac{(\lambda!)^2(\lambda^2-1)!}{(\lambda^2+\lambda-1)!} \tag{6-6}$$

这种体积元可能在某种催化帮助下，将开始把核苷酸序列唯一地翻译成为氨基酸序列，但是仅仅要求体积元中存在着特殊分配的“涨落”。为了稳定这种类型的翻译，我们必须在核苷酸序列中发现那种序列，它们在翻译之后能更好地利用该密码的序列。只有这种核苷酸的集合体才能对于密码和翻译机构（由连接物和活化的酶的特殊集合所构成）提供稳定的、可再产生的信息源。它为了对抗其他竞争系统而被选择，特别是对抗那些不是独特的、并从而总是包括某种无意义复制的系统，它就必须形成在本章的前两段中所描述的自我增强的超循环。①

找到能够促进某种特殊翻译功能的核苷酸集合的概率，与找到实现这种功能的蛋白质集合的概率，都以相同的前提为基础。这些聚核苷酸如果以某种方式被翻译了，那么就代表了多肽的随机序列的一个集合；这样我们就必须从相同的假定出发，即从找到 λ 种特定的“配位者”的假定出发，不过，因为该系统现在必须加强一种特殊的（从 λ! 种可能中来的）独特的分配，所以不再有 λ! 种简并性。如果随机核苷酸序列以相似于随机多肽的浓度水平出现，那么在确定的体积元中找到特定集合的概率，又将由 λ 组的 λ^2 个元素的“反复变更”数的倒数给出。

这一联合概率则基本上由下式给出：

$$P\sim\frac{(\lambda!)^3[(\lambda^2-1)!]^2}{[(\lambda^2+\lambda-1)!]^2} \tag{6-7}$$

由如下例子说明：

---

① 一种可能性是，连接物从一开始就是具有双重功能的、相当伸展了的核苷酸结构：1. 作为携带某个反密码子环的专一性连接物，并由激活酶专一地识别；2. 与其序列一起代表了酶 E_i 的特定信息（I_i），E_i 是该超循环中的成员。不过，在这个阶段得不出关于集结系统复杂性的任何明确结论；可能还包括一些短序列的连接物，这些连接物具有高的丰度。

λ	2	4	8	20
P	2×10^{-2}	1.6×10^{-6}	4×10^{-16}	5×10^{-30}

这个联合概率还包含了一个描述聚核苷酸和多肽序列的相对浓度比的因子。在此涉及到诸如任何序列都有相等的"先验"概率那样的相当不现实的假定，这就限制了把这个公式大致运用于估计某些数字。

重要之处在于，这种处理是要找到多肽随机群体之中某种功能（即氨基酸一密码子分配）的存在概率与再生概率，而不是去找出某些序列巧合的概率。相当大量的不同多肽序列能体现这种功能，此数是如此之大，以至确实在任何随机群体中都有发现它们的好机会（参见福克斯及其同事的实验所表明的，在任何随机合成的一组多肽中类似胰凝乳蛋白酶的功能）。让我们把这个概率记作 p，而不论其特殊形式是什么。然后，把具有相同功能的相同概率，（在翻译之后）应用到（相等浓度的）聚核苷酸的随机序列的群体上。在第 1 种情况下，对于分配我们仍然有 $\lambda!$ 种独特的选择，而第 2 种选择必须与第 1 种选择一致。联合概率从而成为 $\lambda!\,p^2$。所要寻求的，是功能的巧合而不是序列的巧合。如果我们求得找到这样一种核苷酸序列的概率，这种核苷酸在翻译之后准确地类似于多肽序列（它起动这个特定的翻译），那么这个概率将低到 10^{-130}（对于 20 种的 100 个氨基酸而言），这就再一次论证了在第 1 章中讨论的威格纳的论据。

然而，在这种估计中隐藏着很多危险，使得我们在获得更多的关于随机组合成多肽催化作用的实验证据之前，难以进一步深入细节。这个主要的论据，即某种催化专一性不是一种或非常少的几种给定序列的某种独特的功能，而是在规模充分大的任何随机群体中频繁出现的功能，能够由实验加以检验（参见第 7 章）。即使没有这种数据，我们也能够估计出，在什么样的复杂程度情况之下，从随机涨落开始某种独特的翻译变得完全不可能。

关于随机开始翻译的概率之结论如下（乃参见（6-7）式下面的数据）：开始某种二元翻译系统似乎是很容易的，但是两种字

符(或字符类)不足以对功能部位提供充分的专一性。曾有人提出，开始密码进化的最优化程序中，最少需要4种字符。4字符翻译的集结概率仍是一个合理的数量。对于一个8字符密码，它的数值接近于一个极限，这总在地球尺度和早期进化时间范围(这很可能显著少于10^9年，即大约为3×10^{16}秒)以内，它的实际浓度看来是合理的。8字符密码可能是由*AU*系统提供的。而且，8种氨基酸足以构造任何类型的功能上专一的序列。它甚致不必仅从8种氨基酸开始，而从8(或者更少的)类功能上有关的物种开始。这种情形亦适用指令密码，这种密码能够始于简并类，后来可按某种最优化程序继续进化。我们因此可以作出结论，我们今天所发现的特定密码始于某种随机涨落，这看起来并不是根本不可能的，所以我们不必假定某种高度专一的(直接的或间接的)，氨基酸—密码子相互作用。如果实际情形确实如此的话，那么，任何独立的进化系统(例如，在宇宙中的"某个地方"，或者在实验室的"某个时刻")能够利用某种不同的密码，但它仍然是基于类似的原理。而且，现在的密码在起源上也许并不是唯一的一个；不过，普遍性是由激烈的非线性选择程序所保证的。

另一方面，应该强调指出，此时正确说法只能是："我们不可能排除……"所以，上述估价的唯一意义在于，由此去发现在随机开创事件中什么样的复杂程度极不可能的。奥格尔提出了一种有力的论据，他认为翻译系统是逐步地或不断接近地完成的，它以一个或两个优先的连接物—氨基酸分配开始，而不是从集结一整部词典开始。无论何处出现了这种"内在"相关性，它们都会增加自复制功能网络的集结概率。

最后我们可以问一下：我们怎么能从物理学上理解这种翻译的"随机开创"呢？

它也是选择理论的价值标准的结果，这反映在普里戈金、格兰斯多夫的非线性不可逆热力学原理中。无论何时，对于具有由反应系统的某些性质所定义的以及详细说明了外部条件所定义的具有选择行为的自组织系统，(由于涨落或突变出现了较高选择价值的新物种或系统，都将引起不稳定，即先前的定态被打破，并建立起由具有较高选择价值的物种或系统所支配的新定态。新物种的"破土而出"受到一定的限制，并能够用随机理论来正确描述。

我们的结论如下：

核酸提供了内在的自组织前提。不过，它们要求高辨识力的催化活化耦合因子，以便构造出大信息容量。"信息"只有通过功能联系才能获得意义。潜在耦合因子的任何涨落，都将导致独特的翻译，它通过形成催化的超循环而得到加强，并提供了巨大的选择优势并打破先前的与自我复制无关的定态。

作为这种不稳定性的结果，这种功能关联的集结过程（我们可以把它称作生命的起源），原来是一个不可避免的事件。——假若在足够长的时间中保持适宜的自由能流的条件。最初的事件并不是独特的。无论如何，作为非线性竞争的结果密码的通用性必将会产生。

7 进化实验

一个理论模型的价值在于它在多大程度上得到实验检验。一个一般理论的价值在于在多大范围内它指导了研究，并为比较研究而确定明确的、可重复的条件。一个良好的实验，通常是排除不正确的选择，而在可能的选择中作出判断。

关于进化的试管实验仍然很少，这是因为实验工具以及对象即明确定义的分子物种，只是在近年才变得可资利用。西格尔曼和他的小组已经引入了这种巧妙明确、概念简单的模型实验。由于它代表了由本文发展的理论所提示的这种实验，所以将详加讨论。

7.1 Qβ复制酶系统

西格尔曼提出一种关于Qβ的学说时，生物学家同行中最初没有多少人赞同。他的学说认为，噬菌体Qβ运用了一种专一识别Qβ-RNA的专一复制酶。为了排除所有的怀疑，制备了高纯度的并经鉴定的酶，它能够在无细胞的介质中复制出感染病毒RNA。的确是无细胞系统合成了包含着所有指令的RNA，在一

个经典实验中揭示了这个事实。无细胞溶液通过了一系列稀释过程(为了复制，每一步都允许充分长的时间)最后产物中所含的起始自然噬菌体模板少于10^{-15}，即其中甚至连一个拷贝都没有，此样品跟原先的样品一样是感染性的。而且，运用对温度敏感的突变体，排除了RNA分子以外的任何成份是信息源的可能性。这种信息载体只是一条正链，因此，感染拷贝的复制需要一个诱导时期，在这个时期中必须积累起互补的(非感染的)负链。甚至能够由一个单模板拷贝所激起整个群体的复制，从而能够导致均一后代的某个克隆。这点已在实验中证实，实验中，试管里分布着稀释的简单溶液。同步开始合成导致这些试管一致地出现0个、1个或更多个模板拷贝，并严格遵从泊松分布。酶进行识别的区域包括整个分子，其中包括分子的两端，这正如"截断"实验已表明的，一链的任何一半都不被复制酶接受。因为正链和负链都必须由同一种酶来复制，所以可期望关于某链中的互补区域有某种对称性。一链的任何"内部"互补性都将在互补拷贝中作为镜像反映出来，因此在3′的一半和5′的一半中对称地排列。

这个非常有趣的结构问题可以由序列分析来解决，这在苏黎世ETH的魏斯曼的实验室中正在进行(并且已取得部分成果)。西格尔曼的小组也进行了一些序列研究，他们的工作表明，在3′和5′端之间的确有些互补相似性。正链的5′端是：

$$_{ppp}G_pG_pG_pA_pA_pC_pC_p\ldots$$

负链也是在5′端以$_{ppp}G$结尾，并包含一条较长的嘌呤序列。这就意味着，正链的3′端包括了与其5′端互补的区域，这是因为，由于与负链互补，它必定有富裕的尿嘧啶和胞嘧啶，并且也以C结尾。

$Q\beta$系统的发现，既有重要的基础意义，也有重要的实践意

义。各种RNA分子的“无可置疑的结构多样性和精妙性”①，解释了选择力是怎样在核酸与蛋白质的相互作用中起作用，以及是怎样指引前细胞的进化。实际应用可以包括运用特定的识别部位：*a*)与变性的、能够干扰噬菌体感染的非感染*RNA*的结合；或者*b*)作为其他对应的修订系统中的专一RNA合成的诱导物。

7.2 试管中的达尔文进化

利用纯净的 $Q\beta$ 复制酶，导致完成了一系列引人注目的实验。在这些实验中，某种分子物种由“系列转移”而置于选择约束下，从而成为“试管中的进化”。一个典型的实验是从某种标准的反应混合物开始的：

pH7.4 的0.25毫升样品［10^{-1}mol/L Tris HCl 与 $MgCl_2$（2×10^{-2} mol/L）、EDTA(3×10[illegible]M)］，以及 ATP、UTP、GTP 和 CTP 各 200 毫微（^{32}P标记UTP的α-P，所以相应于 1 微克合成的RNA每分钟计数4000次），40微克 $Q\beta$ 复制酶（由 CsCl 和蔗糖离心分离提纯）。关于碱基成分和沉淀分析的程序以及各种酶的活性的测定，详见有关文献。

于是，这个实验由以下内容构成，对病毒RNA保温（这里是对温度敏感的突变体 ts-1）来起动合成，以及在特定的合成周期后把0.02毫升的反应混合物转移到0.25毫升新鲜的标准溶液中实现一系列的稀释。第一步反应是由 0.2 微克的 ts-1RNA引起的，并保持35°C20分钟。然后，保温时间由20分钟（第1～13次转移）减少到15分钟（第14～29次），再减少到10分钟（第30～38

① 引自西格尔曼

次)，7 分钟(第39～52次)，最后为 5 分钟(第53～74次)，就得到了最终产物。在每一次转移时，用0.02毫升混合物来计数，另用0.02毫升注入下一个试管的反应混合物中。在0、8、14、29、37、53和73次转移时，一些产物RNA也取来注入产物中，以进行关于蔗糖沉降分析。产物分析提供了复制RNA分子的如下线索(参见图2-24)：在第 4 次转移后，感染性消失了。直到75次转移以前，RNA模板的分子量或多或少地稳步减少；最后获得的恒定的终点产物，其起初基因组含量已有约83％消失了。在亲代拷贝中存在的3600个核苷酸碱基，只有 550 个保留下来。随着分子量的减少，可观察到 ^{32}P 结合率在增加，所以在74次转移时，每个核苷酸的复合速率是起始合成速率的2.6倍。这已通过饱和条件下核苷酸结合的动力学研究所直接验证。图2-25显示了一个例子：在线性区域，速率与缩短诱导期平行地增长（在诱导

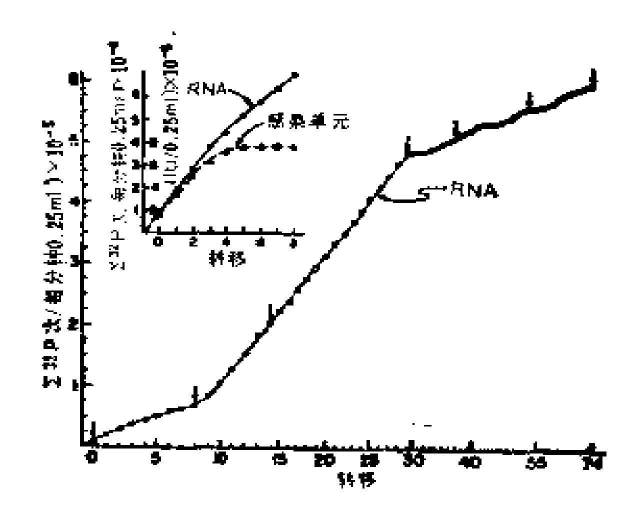


图 2-24 由米尔斯、彼得森和西格尔曼进行的关于 Qβ-RNA 系列转移实验 在正文中叙述了这个实验。在 (0,8,14,29,37,53 和 73 次) 转移上的箭头标明，在此有 1%～10% 的产物被移去并注入反应中以进行蔗糖的沉降分析。保温时间 20 分(第 0～13 次转移)，15 分 (第 14～29 次)，10 分钟(第 30～38 次)，7 分钟(第 39～52 次)，5 分钟(第 53～74 次)。结果表明，在第 4 次转移后生物活性的 RNA 不见了。

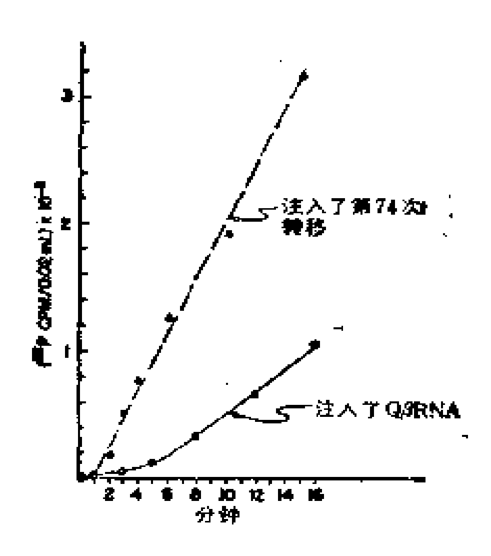


图 2-25 第74次的变异体与原先 ts-Qβ RNA 的合成动力学的比较 两份0.25毫升的标准反应物(细节参见正文)，其中一份注入凝胶纯化的单链变异体 RNA (第74次转移)，另一份中注入了 ts-Qβ RNA (上述两者都是过饱和的)，反应在35℃开始。在所指示的时刻整取0.02毫升，估测 ^{32}P-UTP 的掺入。数据以 cpm/0.02毫升表示。(取自米尔斯等人的结果)

期中，负链的水平必须增加到使正链和负链具有“平衡”比率)。在特定的选择约束条件下，能够观察到速率随着长度向180个核苷酸碱基减少而继续增加。“小怪物”的分数已在拉·乔拉的奥格尔的实验室中得到研究。

初看起来，这些实验结果似乎只反映了对于给定的某种平凡的“进化”反应：“尽快增殖”。从感染的要求中解放出来的*RNA*分子，通过抛弃对快速复制并非必要的一切信息而调整到“极乐”条件。可是，因为识别部位涉及到分子的各部分尤其是两端，所以它们不可能通过在过程结束之前打破循环而压缩复制过程。实际上，压缩转移次数之间的时间间隔并不要求这种行为。反应不会作为转移的结果而中止，已在模板上开始复制过程的酶分子极为可能要继续完成其过程，而不是从模板上脱落下来

去寻求新的拷贝。

事实上，此系列转移实验所提供的东西，是一种近似的"恒反应力"或恒"总组织"条件。

至少就高能单体ATP、UTP、GTP、CTP而论，原先的标准反应混合物的浓度条件，在每一次转移后都得到了恢复，所以，总生成的亲和力平均说来是恒定的。由于压缩了单次转移间的时间周期，便有可能抵偿增长了的复制速率。不过，在西格尔曼的实验中，还存在某种定态漂移，这如图2-24可见。主要是由于巧妙地选取了选择约束，使得在较为定性的情况下证明这些实验是正当的。对恒定条件下保持定态的选择约束进行更精确的测定会影响进化速率，因此要进行定量评价就要进行更精确的测定。

定性结论认为，系统总是偏爱具有最大选择价值的物种。在试管实验的"极乐"条件下，感染性并不是先决条件，而是代表了对快速复制的一种障碍。这个例子清楚地表明，当选择价值总是由速率和识别参数 A、L 和D来确定时，这一切可以由环境条件的变化而发生相当惊人的变化。西格尔曼及其同事在后来的论文中，描述了对试管试验中的各种各样的、已在选择力中适应了不同次级变化的突变体分离问题。在这些能够进入变异体的性质之中，存在着抵制常规的三磷酸核糖核苷模拟物或诸如溴乙烷那样的抑制剂的干扰。通过把酶固定在某种膜上，从而有利于粘性更大的(更长的)链，就可能选择分子量增加了的物种。让其中一种碱基(C)"处于饥饿状态"进行实验，并不产生出缺C突变体；恰恰相反，酶能够调节其结合C的效能，以适应变化了的条件，所有这些实验，都为单链RNA的大量结构、功能可变性和可适应性，及其在早期进化过程中的可能意义，提供着证明材料。

7.3 定量的选择研究

从定量的进化实验中，能够引出更为重要的信息。尽管在这方面有更多的工作要做，但是从迄今所发表的资料数据中，已有可能抽取出一些进一步的信息。

“选择优势”总是与全体物种或系综的复制相联系。如果复制速率依赖于链长，那么，在不增加“内在”复制速率的情况下（与单字符相联系），仅仅失去不必要的信息所导致的链的缩短也可以说是代表了“选择优势”。可是，在饱和条件下，速率也变得与模板浓度无关，于是不存在这种链长相关。“饱和”指的是这样一些条件，它们使得所有的酶分子 E 实际上已被束缚到模板 I_i 上，即 $\sum x_{EI_k} \approx x_E$。这些条件对动力学研究所提供的明显优点，已被西格尔曼在其速率研究中加以利用。如果复制速率的确与模板浓度无关，那么模板分子数（由^{32}P 结合所验证）随时间线性增加（参见图2-25）。如图2-25所示的诱导期的存在，标志了在速率与（或者）正、负链的结合参数中的差别。如果我们把正链和负链分别记作I_+和I_-，浓度记作 x，则包含了图2-25说明所述条件的速率方程可以简单写作：

$$\dot{x}_{I_+} = F_+ x_{EI_-} \qquad (7-1)$$

$$\dot{x}_{I_-} = F_- x_{EI_-}$$

EI_+ 或 EI_- 表示酶模板复合物。一个类似于米凯利斯和门顿的程序，导致

$$x_{EI_+} + x_{EI_-} = x_{E_0} \frac{K_+ x_{I_+} + K_- x_{I_-}}{1 + K_+ x_{I_+} - K_- x_{I_-}} \qquad (7-2)$$

（E_0 指的是酶的总量，包括束缚的和游离的酶。）K_+ 和 K_- 是相

应的酶模板结合成它们的定态类似物的稳定常数。如果在溶液中只注入正链，关于与负链结合的起始斜度不过是：

$$\text{“起始斜度”}\sim F_{-}x_{E} \tag{7-3}$$

在充分积累了大量的负链以后，将趋于某个恒定的“平衡”比 $\bar{x}_{I+}/\bar{x}_{I-}$。

而且，

$$d(\bar{x}_{I+}/\bar{x}_{I-})/dt\approx 0$$

或者

$$\dot{x}_{I+}/\bar{x}_{I+}\approx\dot{x}_{I-}/\bar{x}_{I-} \tag{7-4}$$

运用(7-1)式，可得

$$\bar{x}_{I+}/\bar{x}_{I-}=\sqrt{K_{+}F_{+}/K_{-}F_{-}} \tag{7-5}$$

或者对于把 ^{32}P 结合进正、负链的终态斜度：

$$\text{“终态斜度”}\sim\frac{F_{+}\sqrt{K_{-}F_{-}}+F_{-}\sqrt{K_{+}F_{+}}}{\sqrt{K_{-}F_{-}}+\sqrt{K_{+}F_{+}}}\cdot x_{E0} \tag{7-6}$$

对于 $K_{+}=K_{-}$，则有比例关系

$$\text{“终态斜度”}\sim\sqrt{F_{+}F_{-}}\cdot x_{E0}$$

或者对于$F_{+}=F_{-}\equiv F$有

$$\text{“终态斜度”}\sim Fx_{E0}$$

(参见4.2的结果，不过4.2不涉及到在饱和范围中的酶复制)。

从这些实验数据得出的结论认为，正链的生成比负链的生成快得多(每链都运用互补链作为模板)。斜度的不同，可以用F和K的差异(但不是单独用K值不同)来解释。如果差异只归因于速率参数，则F_{+}可以高达F_{-}的100倍。不过，定量的评价要求更详细的实验证据。终态的变异体(即在74次转移后)，表现了更高的平均速率参数。终态斜度增加到2.6倍（与原先的$Q\beta$-RNA的终态斜度比较而言)。尽管完全有理由相信这种变化大多

是由于速率参数增加所致，但不可能由已知实验数据对单个 F 和 K 的值作出准确评价。诱导期的减少可以归因于速率的增加以及开初就存在正、负链（注意，如果两种链起初就以“平衡”比存在，就不会有诱导期）。现在已有的数据不足以揭示出，斜度增加2.6倍（这肯定意味着平均单字符复制速率的增加）表明了各个变异体 RNA 分子的复制速率增加为 15 倍（与原先的 $Q\beta$-RNA 比较而言）。除了对速率数据的更详细的评价以外，还必须证明分子的总复制时间的减少线性地反映了分子的变短，这不可能是一种以上的酶同时阅读模板造成的，它还有赖于浓度条件。不过，各个分子的复制速率在此有明显的增加，否则，如图2-25所示的，虽然保温期缩短了 4 倍，但是^{32}P 的掺入则不会随转移数而增加。

还要强调指出，上述评价是基于酶—模板相互作用的最简单的可能的定态模型，给出这些评价仅仅是为了表明从定量的速率和机制的研究中获取进一步信息的可能性。

关于选择机制，还应该加上一点说明。人们可从第 2 章的讨论中争辩道，在饱和范围内，选择机制遭到破坏。只要x不与x_i成例，而是恒定的，则此过程就不会是“自催化的”。如果我们指的是某类均一引物的复制，上述说法才能成立。只要存在几种竞争序列 I_i，那么每一种序列都能与酶形成复合体，并由稳定常数标征：

$$K_i = \frac{x_{EI_i}}{x_{I_i} x_E} \tag{7-7}$$

我们得到

$$\sum_k x_{EI_k} = x_{E0} \frac{\sum_k K_k x_{I_k}}{1 + \sum_k K_k x_{I_k}} \tag{7-8}$$

或者

$$x_{EI_i}=x_{E0}\cdot\frac{K_i x_{I_i}}{1+\sum_k K_k x_{I_k}}$$

以及

$$x_{I_i}=\frac{x_{E0}}{1+\sum_k K_k x_{I_k}}\cdot F_i K_i x_{I_i} \tag{7-9}$$

对一致的 $K_i\equiv K$，则归结为

$$x_{I_i}=\frac{F_i x_{E0} x_{I_i}}{x_{I_i}+K^{-1}+\sum_{k\neq i} x_{I_k}} \tag{7-10}$$

物理学解释认为，即使在由$\sum_k K_k x_{I_k}\geqslant 1$ 定义的饱和范围内，某个以小的拷贝数出现的确定突变体 I_i 也是依据某种指数选择机制 ($\dot{x}_{I_i}\sim x_{I_i}$) 被复制的，这直至被选择的物种达到某个浓度水平之前均保持有效，这个浓度水平对应于某个优势：$x_{I_i}\geqslant K^{-1}+\sum_{k\neq i} x_{I_k}$。①

7.4 "独缺"实验

以上面描述的$Q\beta$实验为原型的进化实验，的确能够定量说明在分子水平上的进化过程。然而，只有选好明确定义的、可再产生的反应条件，并应用恒定的约束因素，这些数字才有意义。如果此分子过程的反应模式比较复杂，其中包含了几种同时进化的物种的协同作用，这种机制则可以变得非常复杂，令人望而生畏。

在这种情况下，怎样才能引入试管实验呢？

有一种可能性，我喜欢把它称作"独缺"法。"独缺"指的是一类音乐唱片(众所周知的"独缺音乐")；录制唱片时按正规要求

① 为简单起见，此处探讨了一种"自我识别"机理。互补识别机制的结果原则上与此类似。

要有n个演奏者，但这类唱片只是由$(n-1)$个演奏者录制的。乐曲中的空缺部分要由一个演奏者（在他自己的家中）来演奏，这个演奏者通常是一个热衷于在大型乐队中演奏的业余爱好者。

上述进化实验遵循这个原理（正如西格尔曼的$Q\beta$实验所表明的）。所有的物种中除了一种以外，都是以其最终形式提供的。正是这一种物种要追踪它来自随机前体的进化，而且对于连续的每一个实验都应该改变它（实验的次数须超过存在的物种数，以便还可以对耦合进行分析，因为协同作用并不只是单个过程之和）。这种系统的总进化速率，可以综合所有数据来估计。

显然，这些实验要求一致的，可再产生的反应条件，其中要保持恒定的约束，否则数据将是不可比的。人们可以从想象上构造出一种自动控制的机器，其中由稳定调节（稳定稀释或一定的系列转移）而使单体、聚合物和酶的浓度都保持在恒定水平，并由自动估测^{32}PC以及（或者）用其它标记物的掺入而传达出来。要得以维持单体、RNA聚合物和酶的浓度水平，可借助于微孔过滤材料做成半透壁以制备一个分离系统。在这样的实验中，最有趣的部分——但是困难的——是包含无细胞的蛋白质合成。

我们最后可以得出结论，由这样的试管实验来检验密码起源和分子翻译机构进化的各种各样的模型，看起来并不是不可能的。



8 结论

8.1 理论的限度

这个理论所解释的东西：

是以定态(非线性)热力学理论的稳定性标准为基础，阐释在分子水平上的选择和进化的一般原理。有了一定的、具有特定自催化性质的物质，并保持抵消平稳熵产生所必需的有限(自由)能流，看来进化就是一个不可避免的事件。对于评价实验室的进化实验，该理论提供了一个定量基础。

这个理论可以解释的东西：

是如何去构造简单的分子模型，以代表"活"细胞的可能前体。已考察过4个这种模型，其中只有一个模型能够满足进化到现阶段的细胞生命的所有要求。

这个理论永远解释不了的东西：

是进化的精确的历史过程。"永远解释不了"是所涉及过程的随机本性以及极其多种多样的可能选择的结果。这也适用于关于超过一定时间限度的未来发展。所以："何处不能言语，就在该处缄默。"

8.2 “价值”概念

当我在魏斯曼研究所发表这些演讲时，我的朋友利夫森问我，“新概念通常要带来某个新常数，你的新常数是什么呢？”

为了回答这个问题，让我们区分两类概念。一类我将称之为“新物理学”，另一类是“新的”但是是可导出的“概念”。

迄今我们只经历了两次引入“新物理学”。它揭示了两个基本的自然常数：一个是由量子力学测不准关系表明的普朗克常数，另一个是由于相对论才列入基本自然常数行列的光速。这种引入兴许还有第3次，因为正如海森堡曾开玩笑说过的，我们毕竟是使用 c-g-s 制的。“新物理学”的意义是，抛弃以前公认的、由实验事实所要求的基本原理的普遍有效性，尽管这些实验事实是在清楚明确的条件下获得的，但是它们与新理论的结论不一致。

另一方面，第二类“新概念”并不使迄今公认的任何原理失效；它们仅仅探讨新的方面，并可以从已知原理推导出来。的确还有一些实验事实尚不能够解释，但是，这是由于缺乏洞察或实验，而不是由于违背任何基本原理。玻尔兹曼引入的热力学统计概念，是一极好的例子；他的这个思想被理解为，已知的牛顿力学(后来代之以量子力学)规律应该适用于由分子和原子构成的物质。难题仅在于，粒子的数量很大(例如 10^{24})，而且每一个粒子都需要3个空间坐标和3个动量坐标来说明。引入统计方法带来了重大突破，通过使用诸如温度那样的“平均”量，允许推导分布函数并标征宏观状态（后来由量子理论所隐含的平均规则，证明比在经典理论中更为简单)。我们立刻就会领会

到，这个概念需要引入新的、但却是可推导的量；这个量表达在遍取能量能够分布于其中的所有状态(Z)的平均中，丢失了多少“信息”。这个标征“无知”的量即是熵，其物理意义由玻尔兹曼关系揭示出来，(对于某个微观正则系综)这个关系式可以简单写作

$$S=k\ln Z$$

如果熵描述了用一个(平均)数代表Z个微观态所带来的“信息丧失”，那么，只要“信息”是从Z种具有相等先验概率的可能选择中的一种特定选择[参见(1-2)式]，就能运用相同类型的关系式去描述“信息”。

类似地，如果涉及到具有不同先验概率的微观态时，(归一化的)平均信息量能够用类似于玻尔兹曼H函数的香农公式[(1-4)式]来描述。

在“熵”和“信息”的这种概念定义中的符号不同，从一开始就被认识到了。开尔文的密友泰特，在1868年发表的一篇论文中，表示他不满意克劳修斯为熵选择了正号，他认为熵实际上是“负量”。

在(8-1)式中出现的玻尔兹曼常数并不是一个基本的自然常数。它的物理意义是在温度或热的历史上的概念中形成的。对它可以进行调整，使之吻合于香农的信息概念，这样它就由$1/\ln 2$这种无量纲数来表示了。另一方面，由于熵而引入了一个“新概念”，这在科罗什多里及其他人给出热力学公理基础后变得明显了。

选择理论所讨论的概念具有相似的本质，这就对利夫森提出的问题给出了一种回答。把基本的进化原理理解为在分子水平上的自组织，并不需要“新物理学”，而是需要把宏观现象和基本动力学行为联系起来的可推导的原理。由引入一个价值参

数来表达的此概念，要与信息概念相联系。但是，在系综中的任何特殊状态，而不仅仅是其中某个或某些明确的状态，都能够代表“信息”。引入了与每一种信息状态都相联系的、实际上是连续变化的价值参数，使得我们可以发展起一种其中包括了(“有价值的”)信息的起源或自组织的一般理论，从而把达尔文进化原理与经典的信息论统一起来，把这个概念应用于分子自组织，就为分子生物学提供了一个定量的基础。

选择价值和平均超额生产都是可推导的量，其中包含时间量纲，虽然它们可由引入某种一般的速率常数(k_0)而归结为无量纲形式。

它们的物理意义是什么？

我们考虑一条大分子链，它由一条其中至少有两种单体字符的序列构成。假定可能的序列都具有完全相等的能量（这对于任何现实情形只可能是或多或少有效的近似）。按照热力学标准，所有这些状态都是不可区分的，即“简并的”。某一类长度一致的序列，以及它们分解成为的片段，从单字符的生成是由一致的总亲和力为特征的。不过，如果反应机制包括不同的各个中间物，那么不必要求速率一致。标征每一个体序列复制的“选择价值”的3个唯象参数中，生成速率、分解速率在物理上是由它们的“活化自由能”决定的，品质因子能够与中间态(在此出现指令)的可能反应分支联系起来。在此它们只能通过这3个参数起作用，所有其他的可能(环境的)影响都是次要的。处于定态时，这3个参数的一定结合有赖于特定的约束，并决定着选择价值。

8.3 “耗散”和“信息起源”

普里戈金和格兰斯多夫原理，是选择理论和(不可逆过程)热力学相结合的桥梁。

处于恒流的定态，其特征是熵产生取极小值。如果我们把(内部)熵作为时间的函数；我们就必定会获得一种线性相关(参见图2-26)。例如，平稳地产生的熵，在补偿恒温器中的热流并同时使反应系统保持恒定的内部条件下可以加以测量。支配这种定态的，是“被选择的”序列(或收集)，其相应于

$$\bar{P}=W_{max} \qquad [参见(2-37)式]$$

现在让我们假定一个(随机意义的)涨落，它由比在先被选择的拷贝有较高选择价值的突变体的产生构成。这等价于熵产

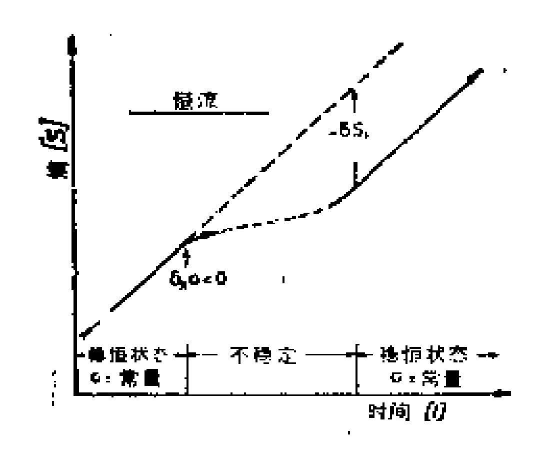


图 2-26 处于恒总率持流的选择过程的熵——时图。显示了某种选择优势W^F突变体>W^F主宰拷贝的某个突变体的出现，相应于由某种不稳定(即定态的打破) 引起的熵产生的负涨落。先前的主宰拷贝要么死亡，要么生长起来处于支配地位。因为两种物种的能量相同，所以两个定态的斜度(代表熵产生 σ) 是相等的。处于定态的两条熵曲线之间的恒定差，反映了由总组织增加所表示的有序增加。

生中的某个负的变化（即平均速率随着简并类的总亲和力减少而增加）。按照普里戈金和格兰斯多夫的原理，这种负涨落必定导致破坏已有的定态，如果外部流保持恒定，则这种定态就不可能保持。所以，在热力学理论中，处于恒流的进化行为是由出现不稳定为特征的。如果我们在重新达到定态以后比较这两种被选择的物种，那么，真正发生的情况是，由有序增加所反映的“有价值的”信息发生了变化。在图2-26中，熵产生的负涨落由曲线的斜度变小来标志；由于不稳定，涨落被放大，因而直至趋近新的定态。因为外部流保持不变，又因为突变体拷贝与其前体具有相同的亲和力，所以达到新的定态时，就恢复到原先的斜度。所形成的S_i的绝对值的恒定差（参见平行的实线和虚线之间的距离），正好等于熵的差，这可归因于由组织程度（即处于已组织的形式的字符分数）所代表的内部有序的增加。这完全不违背第二定律。不过，热力学描述反映不出如下情况：为了产生出新的（更“有价值的”）信息，并由此减少内部熵，几乎一致的群体（或它们的一些重要组分）必须完全可以互换。

如果允许不同序列的自由能不同，从而生成过程和分解过程的亲和力不同，那么这种平衡就变得更复杂了。任何时候，这都将导致选择价值的某种变化，并在进化行为中反映出来。它也将显示在恒流中熵的平衡上，因为此时总的亲和力与各个亲和力不再一致。不过，如果竞争序列的适当约束条件是由聚合体和单体的恒总浓度来表示，而不是由“恒定的平均亲和力”来表示，那么对选择有决定意义的量仍然是选择价值。在这些缓冲条件下，不会出现自由能短缺，于是对于选择来说，任何经济化都没有多少重要性。在此可清楚地看到，选择理论的概念超出了不可逆热力学的概念。在此对于复制有意义的，是各信息内容及其“价值”，而不是非专一的熵平衡。选取上述简并序列

作为例子，是为了表明，作为进化动力的“选择价值”，它是一个与不可逆热力学相联系的新变量，但是它超出了现在的不可逆热力学的范围。特别是关于更复杂的复制系统中利用自由能，还应该作出进一步的说明。

常有人说，进化的指导原理是最经济地利用自由能。这并非是普遍正确的。在自由能供给过量的地方，系统将利用任何选择优势而不顾其“代价”如何。然而，如果能的利用由选择价值的影响而变成次要的约束时，则系统将作出相应的反响。与高“选择价值”相联系，而不是与最经济地消耗自由能相联系的信息利用，是进化中的决定性因素。

在这个方面(即运用于物质的自组织系统)，“信息”不只是一种“结构的功能相关”。它是用来说明在统计平均处理过程中某些细节的丧失。它指的是一个系综中唯象上可区别的状态中的一个状态，在此系综中，总的可能状态数可以大大超过实际上存在(或能够存在)的可能状态数。这样的信息，只有它能够在生成和解体的动力学过程中保存自己，才具有意义。它也许意味着，存在着导致它的在先的信息；但如果我们追溯至它的起源，那么我们就只得说，“原始”信息，代表着它自己复制并变化到由较高多余度所表示的更为“安全”状态的功能。然后它还可以包括次要的“非复制的”信息，这种信息作为有限范围和时间中的与某种功能的指令相关。“选择价值”，表征了信息载体的实施性质，以评价它们的生存变化，并使之保持最稳定的(或最适宜的)状态。

8.4 选择和进化的原理

信息量愈大，分离出下面两个过程就愈正当：

1. 在许多不同状态中的选择。

2. 已选择状态的进化。

如果结构容量是如此之小，以致所有可能的可选择状态都存在，那么两个过程合而为一。然而，可能的配容数通常大大超过能够存在的状态数，而且也只有在这种条件下，信息概念才会是一个有用的概念。例如，在大约10^{10}个核苷酸中贮存的人的基因组的信息量，不过是进化造成的从$4^{10^{10}}$种可能配容(其中当然包括一定的简并)中的一种或几种选择。

在选择与进化的最适条件之间可能存在某些不同，进行优化时需要对它们进行折衷，这正如选择性自身也是在(由于"粘性"结果的)精确性和动态的可变性之间的平衡。

在恒定选择约束下，选择是一个过程；在这个过程中，通过趋近特定环境中信息载体群体里的最高选择价值，使平均生产力得到最优化。

这个过程是由最优化原理来表征的：

$$\bar{P} \text{ 或 } \bar{E} \longrightarrow W_m$$

对于非线性系统，可以代之以更一般的积分关系式(2-34)。W_m表示在竞争者群体中的相对最大值。在一定的约束条件下(最优化程序)，$\bar{E}$ 也趋近于这个值。

选择过程通常包括最经济地利用信息量。在固定的环境中，那些为了包括所有的必要信息所要求的最小字符数的序列（即那些不携带任何无用信息的序列），通常将表现出最高的选择价

值，即A_i(生成速率)、$1/D_i$(寿命)和L_i(精确性参数)都是尽可能大。为了保持已选择状态的信息，需求精确性有最小阈值[例如，由(2-45)式定义的L_{min}]。

对于最优选择，所要求的信息转移精确性必须适应于所要转移的信息量。

另一方面，如果L尽可能的小，则进化是最快的；但是它以选择为基础，所以L必定在阈限L_{min}之上。变异性随L的增加而减小，且当L接近L_{min}时，总是有利于系统进化的。

在由选择标准加上一些约束的条件下，进化代表了进一步最优化的程序。

例如，在某个属于极端的恒总组织约束的恒定价值参数的"线性"系统中，所有相继被选择的物种，其进化途径被限定为W_m的单调增加：

$$W_{m1}<W_{m2}<\cdots<W_{\text{最优}}$$

这就排除了一切通过W_m最小值的进化途径，在其中可以发现几个将导致较高终态价值的W_m。不论这些约束条件是什么，它们都要减少选择数，从而很可能阻止了最好的选择。人们可以得出结论认为，对某个具有大的、不饱和的信息容量的系统，不等式

$$W_{\text{最优}}<W_{max}$$

普遍有效，它表征了一种有限差别，莫诺也许把它称作在"是"和"应该"之间的差别。W_{max}代表绝对的最大值，要趋近它就要求起始时存在或可利用"所有"信息。如果构造一个"信息"空间，其坐标代表了所有可能的"信息"，则系统坐标点的进化变化将描绘出一条基本的非遍历轨迹，这条轨迹指明进化过程的不可避免性。这就定义了适用于所有活系统的特殊时间方向。它与任何不可逆过程的单向熵增加相联系，但比这种熵增加显著得多。

对于非线性系统，在进化期间W_m的绝对值必须增加这种

说法并不普遍正确。出现信息载体之中的内部耦合，与环境的变化是等价的。而且，次要的变化(例如污染)可以引起所有的选择价值普遍减少。进化则仍然由 W_m 序列来表征，每一个 W_m 都表征着一个具有最优行为的物种或集合；不过，这种序列不必是单调的。它也可以由于环境变化而减少，但是随后它将推动此系统向更大信息量方向进化。为在环境变化中生存，或为了取得它自己的优势而改变环境，则需要更多的信息。所以，

进化可以包括提高选择价值和利用更大的信息量。

至少在分子自组织的更高水平上，第二种影响将是主要的，这是因为在复杂系统中的耦合的可能性变得如此巨大，以至大量的突变体可以引起价值普遍减少；只有出现了适应于改变了的环境的突变体，才能补偿这种减少。只要单个信息载体开始相互作用，“好”和“坏”这些术语就取得了意义，从而相互引起它们的“价值”增加或减少。

在分子水平上的进化可以看作一种博奕，追求随机出现的事件中的优势这种选择“本能”取代了其中的有智慧的博奕者。所以，由冯·诺意曼引入的、近年来已发展到很高水平的博奕论，是进化理论进一步推广的关键。①

8.5 “非决定”但“不可避免”

“选择”和“进化”有些类似于②热力学中的“平衡”，它可

① 分子生物学中最优化问题的特殊数学处理方法是由赖欣伯提出来的。从博奕论的观点来处理一般的进化现象，是由卢旺廷和沃特伯顿提出来的。木村资生和马凯强调了“信息”可以起源于“噪声”这个事实。

② 注意区别：a)“最优化”和“最大化”；b)定态缺乏微观可逆性和 c)全微分的适用限度。

以用极值原理来描述，这个事实使我们能给达尔文原理以物理学基础并建立定量表达式。这样做，达尔文原理就不仅仅是指历史途径，而且是指一种可从物理学推出的规律，这种物理学支配着物质自组织的一般过程。不过，较高级的生命形式尤其是“智慧的”控制形式的进化，还将要求进一步考虑一些其他原理。所以，我们这里把我们自己限制在分子生物学的领域，即限制在分子水平上的自组织过程。

我们给“价值”概念以物理学评价这一事实，可以修正我们对达尔文的解释。在现代生物学中，有一种被广泛接受的关于达尔文原理表述的看法，用沃丁顿的话来说：这种表述只不过是“一种陈词滥调或同义反复”。斯滕特在其《黄金时代的到来》一书中写道：“众所周知，适者生存不过是同义反复，即生存者生存，所以在这里‘不适者’不代表一种客观的科学判断，而是代表一种主观的价值判断。”

只有“价值”(我们亦可把它称作别的什么东西)仅代表了另外一种完全不确定事件的结果时，上述说法才是正确的。因此，重要之处在于，把由(决定论的)唯象理论所引入的价值概念，放进随机分析之中。必须对“选择”(它在给定的信息量之中)及进化（它利用了选择程序以便于趋向某个最优价值）分别进行这种研究。在第3章中，已讨论了随机分析的基本概念。在那里所获得的其中一个结果，对于上述问题具有重大意义。它处理了n个不同的、但都是W_i简并序列的精确复制($L=1$)。随机分析表明，这种系统由于它的自催化行为，它总会（或者高度可能）把信息量集中到一种序列上，不过这个序列将聚集起n个拷贝的多余度。这代表了一种真正的“适者生存”情况，因为在此无法预见**哪一个**模板将生存，这些模板都是物理学上不可区分的；它正是一种被放大的涨落，即是作为内在的复制机制的一

种结果。

不过，因为L决不可能正好等于1，所以这个例子代表了一种不现实的特例。如果完全简并以及仍然保持产生与分解之间总平衡的条件，那么每一个体序列的选择价值都必定为负($F_i=D_i$，但$L_i<1$)，因此在完全简并的信息载体中决不可能出现稳定信息的选择。只有在不同信息载体的选择价值表现了一定的分布时，才会在定态出现具有最大价值W_i的物种的稳定选择。

这种过程也不是完全决定的，但是对于任何自我调节的随机过程，涨落极限将随已选择的拷贝数的增加而减小。因为选择通常始于为数不多的、或甚至是一种突变体拷贝，所以涨落现象具有重大意义。

与上面讨论的完全简并问题密切联系的，是"随机漂移"现象。当由"中性"突变产生出几种其选择价值是简并的物种时，这种现象就出现了。在文献中，往往把它称作"非达尔文"进化。以不同系统发生水平上的物种的蛋白质序列分析为基础的、关于进化速率估价的早期工作中，忽视了这种现象的出现，但把它称为"非达尔文的"就有些不合适了。"中性"突变体及其"随机漂移"，当然也完全属于本文所勾画的更为抽象的选择概念的范围。

由复制过程中的"错误复制"，或者由作为进化的最优化程序基础的其他种类的突变体，引入了对于决定性的更为严格的限制。虽然对于突变来说，不同的字符位置并不完全等价，并且所形成的突变体仍然与其主宰拷贝相联系。但是，如果在突变的"因"和"果"之间绝无联系，那么就不会有什么突变，从而整个过程看起来是随机的。而且，导致突变的基本物理学过程由于其量子力学本性，内在地是非决定的。强加上自催化选择

过程，可以把高选择价值的突变体过滤出来并加以放大，从而就价值原理而论是减少了非决定性。不过，各个拷贝选择仍然存在非决定性。结果造成不可能追溯精确的历史途径，或不可能预见超出了一定时间限度的准确的未来发展路线。而且，正是非决定性使达到绝对的最高价值成为不可能。

在较高水平的、尤其在群体遗传学的进化中，研究选择的最有效的途径是从生存这种事实开始的。这正是这种研究的成功之处(与费希尔，霍尔丹和赖特的名字相联系)①，这导致了一些生物学家把达尔文选择原理作了同义反复的解释。②

可是，如果我们能够把生存与物理学上的客观“价值”联系起来(这是速率和相互作用参数的相当特殊的结合)，那么选择原理就不再代表一种无足轻重的同义反复或陈词滥调。任何原理，在它的逻辑内容得到阐明以后就可以多少清楚一些了，因为逻辑将揭示出是同义反复还是对应关系。

进一步的结论是，如果生命进化是以可推导的物理学原理为基础的，那么，尽管它的途径是非决定的，但是它必须被看作一个不可避免的过程(参见后面)。在第4章到第6章处理的模型以及在第4、第6和第7章讨论的实验表明，它不仅仅是“原则上”不可避免的，而且是现实的时间间隔中充分可能的。它要求适宜的环境条件（这不是在任何地方都具备的)，并保持这些条件。这些条件曾在地球上存在过，并必定在宇宙中的许多行星上继续存在着。只要能够供给能量，这种进化过程就会继续下去而没有时间的限制。这样一来，任何关于“固有的”时间限度的预言，终将取决于我们关于宇宙能源的利用性知

① “选择价值”这个术语出自他们的著作。

② 沃尔特拉和洛特卡的著作值得一提。它联系着对特定的竞争生长和“斗争”问题形式的数学处理，但与大分子中“信息”发生这个一般问题没有多少联系。

识，因而这就密切联系着宇宙学问题（关于宇宙学迄今尚无定论）。

本文中关于选择和进化原理的抽象公式，当然没有包含如下假定：进化实际上发生在定态的极端的.抽象的约束条件下。本文已经强调了与热力学平衡的相似之处。对涉及到蒸汽机最大(效率)的卡诺循环的抽象处理，带来了对(平衡)热力学原理的理解，当然没有任何一个蒸汽机是处于、或曾经是处于平衡条件下工作的。如果总是保持极端的约束，进化中的进步将是极其缓慢的(并且有许多社会学的含义)。不过，这些条件揭示了原则性的东西，使我们可进行模型分析，并告诉我们如何进行重复实验，从而最终引导我们去重构某些进化事件。

我们接下去谈最后一个问题。

8.6 生命现象能够用现在的物理学概念加以解释吗?

如果简单地回答"能够"，那就把我们置于一种困难的境地，因为这就要求我们证明这种回答，例如要用完全归纳法来证明这种回答。聪明一点是把问题倒过来，即无论何时如果有人说物理学对生命不提供任何解释，那末就请他证明他的结论，或者更好的是，我们只须用一个反例就可以否证他的结论。

我们有这种反例吗？这大大有赖于我们对生命定义是否有一致的看法。

奥巴林曾提出如下的性质来作为"生命"一词的定义基础：

代谢，

自我复制，

突变性。

象冯·诺意曼的“自我复制的自动机”那样的机器，就能够满足这种定义。这种机器肯定会有“代谢”。在电插头或油箱的环境中，它能够找到足够的“食物”，并进行任何一类工作。当然，这种自动机的特征不仅仅在于具备按照某个程序而复制出它自身的能力，而且在于复制出此程序的能力，从而使得任何后代机器能够复制错误，这些错误可能会导致有优势的“基因型”突变体。我们会认为这种机器人是活的吗？大概不行。因为它需要人去启动它，所以我们会把它称作是“人造的”。

在第4到第6章中处理的复制性大分子循环，是另一个例子。这些循环能够自我启动，但是，任何复杂程度小于第6章处理的催化超循环的东西，都不能看作是“活的”；这种催化超循环的特征有10个(参见6.2)，其中包括了奥巴林提出的3个。这种循环的存在只取决于：

a) 如同在第4到第6章详细提到的物质的某些化学性质，至少在原则上能够用量子力学理论来解释；以及

b) 要有一定的物理条件，我们完全有理由假定地球上曾存在这种条件。

我们必然得出这样的结论：生物学基础不需要“新物理学”，但是同时我们要认识到，我们从这个结论还没有得到多少东西。从单个大分子到催化超循环或者“活”细胞的这一步，肯定没有从单细胞向具有自我意识的、智慧的人类之过渡那样惊人。要理解在这个过渡中包含的各种各样步骤，很可能不需要多少“新物理学”，但是很可能需要如同第一步所要求的许多进一步的(可推导的)“概念”。

正如维特根斯坦在50年前说过：

“人们只是在生命问题消失时，才看到了生命问题的解决。”



9 总结

现有的工作是些评述性文章和原始素材方面的报告。这些工作转到物理学家和生物学家方面。对于物理学家来说，他们必须对一些生物学家或者生物化学家完全熟悉的事实作出总结，以说明从那些前题出发的生命现象理论已经建立起来了，并且说明为什么是完全肯定的，以及为什么不讨论别的模型。同时还应当向生物学家指出，生命过程是被物理学原理支配着的，并且不能回避定量化的公式。一些对物理学家来说是熟悉的术语和解决方法，通常还有半定量近似的清晰性从而被详尽地讨论。

问题的焦点是：

"生物学是否能够通过物理学以它现代的方式建立起来？"

具体答复至今仍被概括为一句话，应当说，对于至今已被充分研究过的生物学的过程和现象没有任何答案。因为我们所熟悉的形式的物理学还做不到这一点。如同在无生命世界中的宏观现象，尽管对现象作详细描写的限制不在原则方面，而是在复杂性方面，但是并不排除我们熟悉的物理学原理对于生命现象以一种特殊的形式表达出来。首先要说的是对于信息产生的理论，其特征方面和物理学可以引出的东西——一盘的概念支

配着进化的最优化过程。或者说，这些过程的暂时的优先取向起源于不可逆过程热力学的稳定性临界理论，而进化则在原则上是不可避免的。

下面进行详细的陈述：

进化论原理的唯象公式在第 2 章进行了表述。达尔文原理是以一个和最优化原则相联系的确定的物理假设出现的，一般并不作为一种和生物圈相联系的不可逆现象。它通过普里戈金和格兰斯多夫的稳定性临界理论和定态热力学理论相联系。有些概念，如选择压力和选择值等可以通过承认所限定的动力学条件(例如，“流”和“力”这些常数)进行定量的和客观的物理描写。这里表达的量的概念为信息论提供了基础。“信息”在这里是指在动力学理论中所描写的分子的一种特征，并且通过自身的复制能力进行评价。它能够使不同能量的状态互相区别，并对一种按照纯粹功能观点进行的“首次”自组织作上标记。

第三章研究了唯象理论。由于单一事件的不确定性所受到的限制，随机理论在大量粒子数的情况下提供了一种可能性，去建立唯象理论的平均行为。它并且进一步导出了唯象理论的这种表述。在这里，进化一概起源于单个事件，通过生长得到“加强”，并在宏观上对此作了描绘。一种在本质上比较强的不确定性，在与封闭系统相比较、与现存系统相平衡中产生出来。这种不确定性不仅表现在个体结构上并且表现在其历史过程中。当然，这种理论和生物学家中传播的个体不确定性相联系，它将消除一种偏见，即“适者生存”的达尔文原理是一种平庸的同义反复：“生存者生存”。假如“最适者”只是通过生存的事实作为一种偶然的结果被肯定下来，那么这种解释也许是合理的。这种情况是少见的，并且在自然界是几乎实现不了的。假定这种情况有一次以一种完全精确的、完美无缺的重复出现，那么另

一方面所有选择量，即竞争着的种的动力学特性的不确定性就会完全退化。这样一来，由于自动催化复制机制，某种偶然的涨落可能得到加强，以致出现这样一种选择，人们不能以任何方式预言其结果。但是所有日常进行着的复制过程，由于相互作用能的限制，都附带有一种已知的缺损。在这方面，它们一般在选择量上相区别。而选择总是通过最优过程来识别的，在那里，"最适者"这个概念是一种具有附加条件的极大值，即所谓最优值。涨落起着一种重要的作用，在选择一种合适的有益变异时，其重要性首先表现在一种唯一的复本形式上。涨落的作用越强，择优选择的"选择优势"越小。

第4到第5章探讨了具体的反应模型。应当指出，"生命力"结构的出现(单细胞微生物的前身)是和选择机制的特殊条件相联系的。它既不由核酸也不完全由蛋白质所构成。以选择机制为基础的复制过程一定是非线性的。而不仅核酸还有蛋白质都能按"准线性"复制机制复制出来。从而出现了这样的系统：核酸所含的信息太少，蛋白质所含的信息太多。信息太少意味着不同的相互竞争的序列不存在；为编码择优功能可复制需收集足够的信息。太多的信息意味着自身择优变异的概率太小。反之，如果一个系统不仅含有蛋白质而且含有核酸，那么可以利用两者的功能方面的优点得到了一种稳定的选择。其优点是：

a) 核酸具有固有的自适应性。借助于此，系统不仅可以复制，而且可以改变一次性建立起来的信息状态。

b) 蛋白质具有巨大的功能方面的能力(特征识别、催化作用和有规则性)。它在建立有序的功能统一的过程中把单个反映步骤联接起来是必不可少的。

由核酸和蛋白质之间的耦合(非线性的)所产生的反应循环等级已经表明了生命系统的本质特征，并且一直进化到活细胞。

出现一种这样的自复制超循环系统取决于一种单义编码系统的建立。第6章后半部分讨论了具有单义有序编码形成的假定。

第7章最后描述了进化论方面的实验。例如西格尔曼首先用$Q\beta$-噬菌体来完成这些实验。第2和第4章发展起来的理论提供了进行可重复测量和定量评价的基础。

现将前述工作的结果汇总如下：

1. 对于核酸和蛋白质复制机制的详尽分析并没有显示出某种根据，从而承认存在生命现象所特有的力或相互作用。以生命系统进化为特征的选择行为已经作为一种特殊的反应系统所固有的物质属性。

2. 每一个通过进化或者选择所获得的系统，其个体的构造是不确定的，其进化的发生过程是必然的，因此这是一种规律。具有某种优势的进化的出现和一种不稳定性相符合，这一点可以用普里戈金和格兰斯多夫关于稳定性的不可逆热力学过程进行解释。因此进化是一种最优化过程，在原则上是不可避免的。而个别进化途径的选择倒是不确定的。

3. 最后应当指出，生命的出现是和一种固有属性的序列联系在一起的，在物理上是被完全单义地建立起来的。建立这种固有属性的前提估计已被逐步实现了，所以“生命起源”或者物种进化都不能描写成一次性的创造活动。

鸣谢

本文所表达的一个个思想，其中许多不能说是新奇的。不过，本文是为物理学家和生物学家写的，其中一部分人有时看起来是平凡的东西，而另一部分人则可能不这样看。请原谅我的文章这么长，是因为我认为整体比一个个思想的总和所代表的东西要多。

关于选择理论的思想，是在与克里克共用早餐的讨论中激

发起的。西格尔曼的进化实验以及普里戈金对“不可逆热力学”的基本贡献对此提供了进一步的推动。目前维也纳大学的舒斯特进行了(且正在进行)关于自组织催化循环的大量计算机计算工作。我们多次有趣地与计算机打赌，猜测进化竞争的结果，结果总是计算机获胜。罗杰·索尼里对我的英文稿进行了校正，还有人耐心地听取了这一切。

谨向他们所有人致谢！

生物信息的起源

艾 根

1 问题和伪问题

在我们这个时代，物理学家必定会把生命现象包含在关于自然的概念之中。

生命是什么？

30年前，薛定谔在他的著名文章中，卓有胆识地冒险回答了这个问题。在生命的分子密码秘密已被揭示的今天，生物学家似乎更谨慎了，用克里克的话来说："众所周知，难以定义生命这个词。"

物理学家们经常碰到这种困难，特别是在涉及到十分基本的术语方面。牛顿在其《原理》中对"时间是什么"这个问题，给出了一个非常明确的回答。但是范曼在他的《物理学讲义》中，竟然给出另一个定义："无论如何，真正的问题不在于我们如何定义时间，而在于我们如何测量时间。"

我们能够测量生命现象吗？或者最好这样提问：我们能够规定生命特有的标准并通过试管实验验证这些标准的有效性吗？

我们必须意识到，生命有许多层次——它最终可归结到物理学家关于"自然"的概念，这正是我们这本专题论文集的主题之一。

为了回答上述问题，我们当然不得不从某些较低的层次入手。我们的兴趣将集中在从非生命到生命的过渡，即“生物信息的起源”问题。有许多问题是属于这个学术主题的，其中最普通的问题是“先有鸡还是先有蛋”。这个问题：在现代生物学中它完全可以用“先有蛋白质还是先有核酸”，或者用更抽象的措词“先有功能还是先有信息”这类问题来代替。然而，如果“先”是指因果关系而不是指时间关系，那么“谁在先”就会显得毫无意义。原因在于，倘若功能为了进化到目前已知的复杂程度需要以信息来表达，那么反过来看，这种信息又只有通过它编码的功能才获得其全部意义。的确，在所有的生命物体中，甚至在最原始的生命物体中，蛋白质和核酸都联接在由一些反馈环路组成的复杂的分级结构中（参见图3-4）。因果关系的开端就象闭合循环线的开头与末端一样没有什么意义。

以“普通的”化学制备为基础进行的实验室实验告诉我们，在前生物条件下，这两类生物大分子可能已互相独立地产生出来。这样一来，如果我们重新表述“先有鸡还是先有蛋”问题，那么与生命起源相联系的时间上的“先”便是有意义的。例如，这两类大分子中的任何一类，是否具备足够的实施力量和立法力量，以至能够自发地启动自组织过程及进化过程？或者换一种说法：它是蛋白质和核酸之间的相互通讯吗？这种相互通讯就是指密码与翻译机的集结，它起动一种收敛的、功能性的自组织过程——这种过程与发散性的、无组织的大分子合成过程不同。

隐藏在所有这些问题背后的主要难题，是生物的复杂性。这种复杂性在大分子化学层次上就已经表现出来了。

2 生物的复杂性和特殊性

考察一下一种典型的小蛋白质分子肌红蛋白，其结构如图1所示。其分子量约为10^4，这意味着多肽链约由10^2个氨基酸残基构成。给定20种自然的氨基酸，我们会发现在这种长度上有20^{100}或10^{130}那么多种可选择序列。

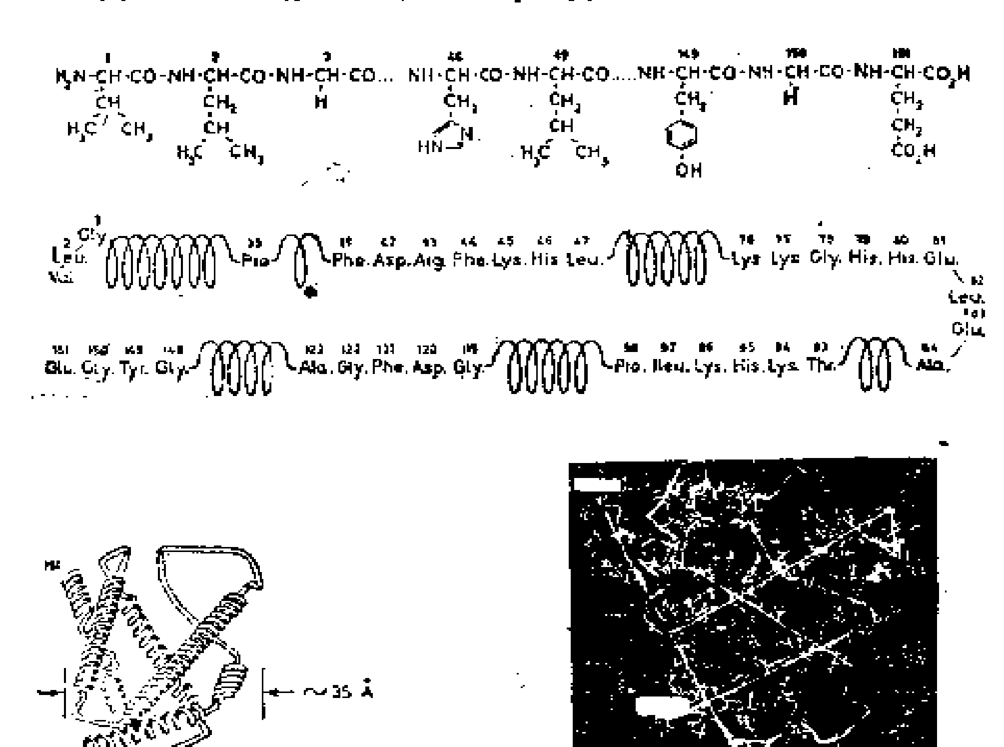


图 3-1　蛋白质的结构层次

(*a*)一级：氨基酸(大约20种)以共价键结合在一起形成线性多肽链。

(*b*)二级：肽键增强了线性链中的CO—和NH-基团的极性。"螺旋"或"折叠"结构是由这两种基团间形成氢键所致。

(*c*)三级：由侧链相互作用而稳定化的空间折叠，功能团之间的距离有一种精确配合的相互关系，从而形成"活性部位"。例如，如图中所示由X射线衍射测定的肌红蛋白的结构。

(*d*)四级：几条多肽链可以结合成包含有调节和催化能力的复合体。通过调节位置处效应物结合上的构象变化，来诱导酶功能的控制。

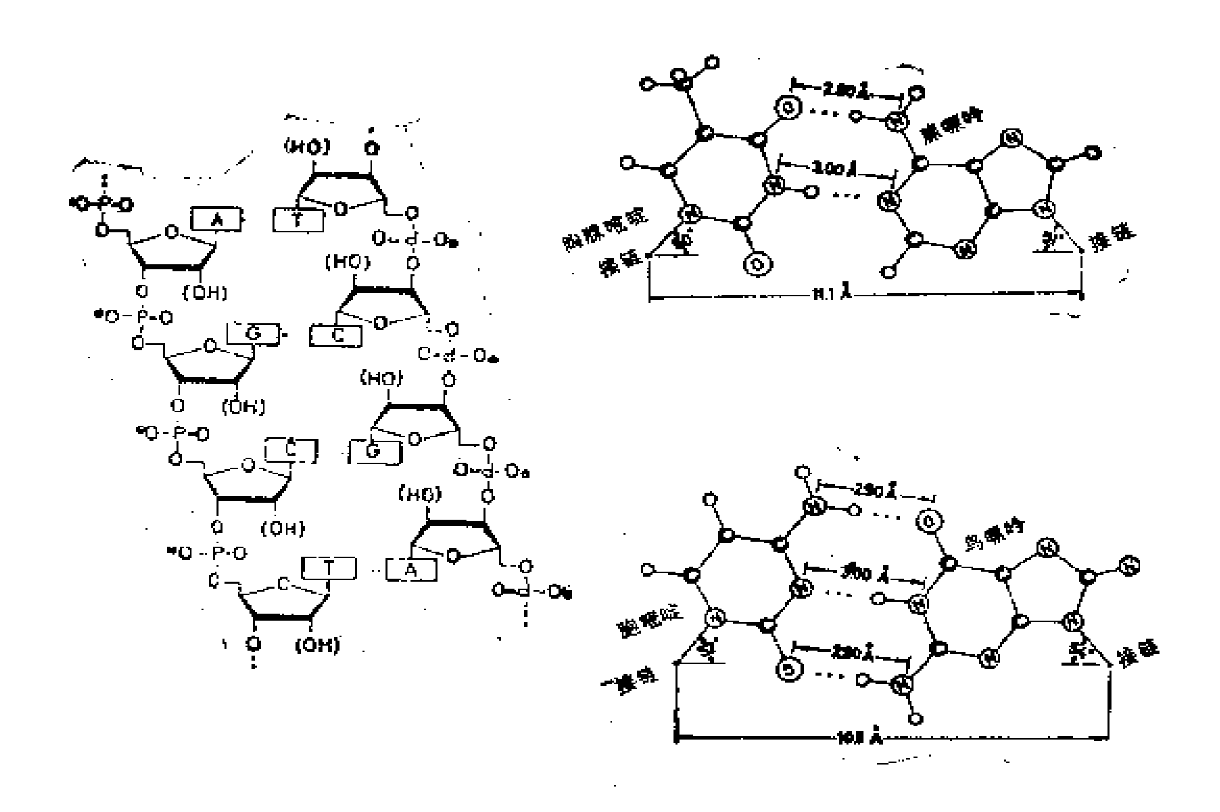


图 3-2　核酸的双螺旋结构(沃森和克里克)

聚核苷酸链的骨架代表了磷酸酯和糖(核糖或脱氧核糖)交替的一种序列，磷酸脱氧核糖通过磷酸二酯键共价结合。每一种糖的亚单元载有A、T、G、C(在DNA中)或A、U、G、C(在RNA中)四种碱基中的一种，这些碱基即遗传字母表中的字符。互补性是建立在下述基础上的，即腺嘌呤(A)和鸟嘌呤(G)分别与胸腺嘧啶(T，或尿嘧啶U)和胞嘧啶(C)的优先相互作用，产生出两对(几乎)一致的几何形状(被聚合酶识别)。

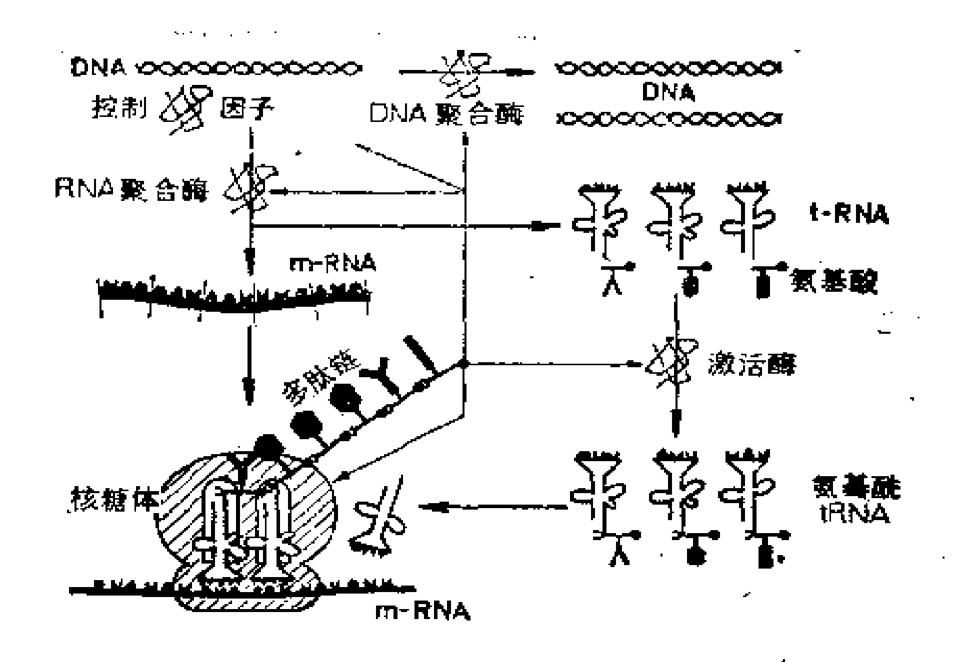


图 3-3 核酸和蛋白质在其复制循环中密切相关

重要的功能联系是：

(*a*)DNA 和 DNA 聚合酶，贮存并繁殖遗传信息。

(*b*)信使 RNA 和 RNA 聚合酶（与作为启动基因或阻遏基因的蛋白质控制因子相互作用），使处理过程能获得信息。

(*c*)转移 RNA、氨基酸及其特定的激活酶(氨基酰合成酶)，充当引起转译的媒介，它们把每一个氨基酸联接到其特定的(带有反密码子的)连接物上。

(*d*)RNA 和蛋白质亚单元构成的复合体——核糖体，结合并传递信使 RNA，与氨基酰转移 RNA 匹配，并催化多肽链的合成。

对于这种为活细胞所运用的自复制的生物合成循环，分子生物学已对其复杂的分级组织有了深刻的见解。

实际上，肌红蛋白是一种我们在自然界发现的比较小的蛋白质分子。大多数蛋白质都是相当大的，在一条多肽链中常可包含300或400个氨基酸残基。而且，这样的肽链还可以聚集成为分子量极大的复合体，从而把它们的催化能力与控制和调节功能结合起来(例如，“变构”酶)。

必须把这些极其巨大的数目与我们通常在物理学世界中遇到的数字联系起来看，例如与阿佛伽德罗常数加以比较；甚至与有关宇宙的物质总量加以比较，这个总量以质子质量表示，其大小也“仅仅”达到10^{78}的数量级。如果我们继续推进到整个细胞例如微生物的水平，而更不用说推进到更高级的组织，这种不相称就已几乎令人沮丧了。

构成一个细菌细胞(例如大肠杆菌)的整个基因组的*DNA*分子，代表了从$10^{1000000}$个可选择序列中的一种选择或寥寥可数的几种选择。假若我们宇宙中化学上的有关状态仅仅存在了10^{17}秒，那么显然，在此类全部可能性中，只有极其微小的一部分有可能已被自然界验证过。(注意：甚至借助高度适应的酶，一个蛋白质分子的合成或降解所需的时间至少需要从1/10秒钟到几秒钟或几分钟。)

复杂性总是被看作是生物组织的必不可少的条件。薛定谔在谈到它时用了“非周期晶体”这个术语。其实要紧的是一种提供信息的品质而不是有形的品质。

分子尺度上的复杂性，是由微观状态的个数来标征的，这个数字极为巨大，而且其中一些状态可以具有特殊的宏观意义。这种数量大大超过了存在(或能够存在)于一定的现实时空限度内的任何状态数。总之，在统计力学的经典问题中，也包含着如此巨大数量的、远远超出任何群体容量的微观状态。不过，它们通常涉及到遍历性系统。这些系统性质允许适当的平均，由此导致了在有限的涨落范围内对这种宏观行为进行决定论的描述。此类系统的再现时间是有限的，纵然这种时间也许远远超过了宇宙的年龄。

另一方面，我们在此将要讨论的系统，涉及到非遍历性。某些独特的微观状态可以强烈地影响宏观行为。在出现这种情况

时，这种“涨落”可被放大，并可以打破先前的稳定态(即亚稳态)。事实上，进化可被看作这类单向事件的一个序列。

生物的特殊性问题与生物的复杂性问题互为补充。关于上述的巨大复杂性，自然界的任何选择能特殊到怎样的程度呢？

由香农、维纳、柯尔莫哥洛夫和其他人发展起来的信息论，涉及到从一个可选择的集合中进行某一种特殊选择方面的概率。让我们考虑某条编码遗传信息的核苷酸序列。来自随机性的偏差，或许可归因于单核苷酸有不同的先验概率以及由于优先结合，这种优先结合类似于语词和句子，它是由语言结构所规定的、或者由给定信息的意义所导出的。对任何k组符号结合，所有这些约束都能够用概率分布来表达。

正如布里渊在他的专著《科学和信息论》中已经清楚指出的，概率论并不解释一条信息的“意义”或“价值”。然而，在生物特殊性背后的是信息的某种语义方面。这种语义方面的最终解释，必定只能在功能的即动力学的性质和关系中找到，而不是在某种丰度和核酸的碱基对或者蛋白质的氨基酸残基的最近(和非最近)的相邻频率中(尽管在功能方面后者肯定是重要的，犹如通信和语义表达需要一致的语言结构)找到。

让我们就功能的独特性表达的生物信息的特殊性多谈一点。对于任一生物实体，比如说一个酶分子，一个细胞器，一个单细胞，或者甚至某一生物体的整体，如果我们分析其动力学行为，我们总会遇到其中一切基本步骤的完善的相互关系。“完善”一词系指，譬如，可测速率处于或接近于被物理学定律所规定的上限。任何化学转化的速率，都受到相遇反应物的碰撞或相应频率的限制。如果形成了络合物，那么有辨别力的识别是建立在一定大小的稳定常数(即相互反应的自由能)的基础上的分析表明，酶最能适合这些要求。另一个例子是一个有成千上万

个蛋白质分子及其他分子的整个细菌细胞的复制。大肠杆菌细胞的一个世代的时间，共计约为20分钟。其中包括对编码在一条约有400万个核苷酸的线性链上的信息进行顺序阅读、翻译和处理。一条400万个字母的信息，相当于一本排印得很密集的厚达千余页的书。按共计 20 分钟计，则处理每一字母的时间小于1/1000秒。关于这种密码阅读和复制的一些基本化学步骤的速率，已经直接测得。有辨别力的识别单码字母，所需时间少于10^{-6}秒。

显然，至少在分子水平上，遗留下来需进一步改进的东西不多了。看来千千万万个基本化学过程中的功能性的相互关系几乎是完善的。这种完善标志了生物结构的真正的特性吗？有多少种选择对象实际上被验证过？

一个包含λ种ν个残基的序列，在K个位上有取代，则可选择的拷贝数是$\binom{\nu}{K}(\lambda-1)^k$。于是一个蛋白质小分子（λ=20，ν=100，参见前面）只有 1900 种"单误差"拷贝，但是，若 K=9，即对该原序列中不到10%的位置进行取代，则可选择的突变体数就已达到阿佛伽德罗常数的数量级。因此，除非运用某种等级原理，否则对可选择序列的检验是十分有限的。

我们可以提出三种不同的可能解释，来概括这种关于生物特殊性的讨论：

(1) 拷贝选择只受物理力支配。生命实体的出现，只不过是这种物理力相互作用的表现形式。

结果：生命独立地出现，例如无论是从另一行星还是从试管中独立出现，都会导致一致的结构，特别是导致相同的密码和酶机构。

(2) 选择是"特殊的"事件（或最好说是巧合）。它出现的期

望值低得可以忽略，但是相互作用的随机本质并不排除这种意外的、罕见的涨落。
结果：在宇宙中任何其他物理和化学的条件类似于早期地球的地方，发现生命独立起源的概率将小得可以忽略。

(3) 在适当的空间与时间范围内以及有利的物理条件下，自复制及进一步的进化系统的集合过程，以一定的期望值出现。此值介于大分子顺序的任何分布之中。这种顺序基本上是随机的。此类大分子系指蛋白质与核酸。因此起初的拷贝选择是偶然的，但是随后进化优化到特殊效率的水平，这种优化是受物理学原理支配的。
结果：无论什么地方，只要具备合适的物理和化学条件(即，正如天体物理学家告诉我们的，这在宇宙中并不少见)，就应该发现生命。不过，个体结构，尤其是分子密码和机构，应与我们所知的系统大不一样。因此那些结构或许不可能被地球上的任何物种直接利用，尽管它们可以是以类似原理为基础构造的。特别是，在光学活性结构中发现不同手性的机会应该是50%。

我已经列举出这三种假说，因为它们导致截然不同的结果，这些结果兴许有一天会被空间实验、或者可能更快一些被试管试验加以核实。

第一个假说的困难，尤其与基因型层次上的现象及表现型层次上的现象之间的独立性相联系。我们必须承认：我们不知道任何简单的结构—功能关系，这种关系可能会提示一种由内在力支配的对进化的独特指导。这样一来，我们将不得不运用一些迄今尚不理解的(也可能是新的)物理学概念。

第二个假说，将是任何理论都力所不能及的。我们所需要的不仅仅是某一事件，而是整个事件系列，其中每一事件的期

望值都是(几乎是)零。

第三个假说的吸引人之处在于，它涉及到开始拷贝选择的巨大多重性。例如，在任何随机的球蛋白结构中固定几个点，就足以规定某种确定的催化优势。或许会找到1/1,000,000，即在10^{130}个可能的序列中有10^{124}个序列。(当然，任何确定的数都取决于我们所谓的“优势”是什么。)再者，特定的起始拷贝期望值还是小得可以忽略。仅仅是某一起始拷贝的出现这样一个事实具有充分大的概率。相继的进化优化过程具有足够多的分支点，从而完全不能确定任何特定途径；但是另一方面，优化可能是某个物理规律的结果，就此而言前述情况也不一定属实。最终结果将是某种独特的结构，例如，具有优化空间坐标的K组确定位置。可选择的最终目标(从而又有可选择的最优结构)的数目，肯定将大于1(即10^x，$1<x\leqslant 130$)。例如，如果x等于5或10，那么，任何这种最优结构仍将是极其大量(10^{125}或10^{120})效率较低的结构中的一种。

这与诸如下棋那样的策略性博奕有一定的相似之处。开始的几步具有走对的好机会，因为大部分的下子有继续进化的潜势。在中间阶段，引入了大量的分支，而在结束阶段，可能的解决方法被压缩到相当少数几个等效的选择上，这种选择具有最佳的“将”棋。

目前已获得的带有随机结构的实验证据还是支离破碎的，它有利于中间型假说，而不是两种极端中的任何一个假说。然而，相当多的物理学家和哲学家似乎专注于第一个“上帝不掷骰子”的假说，而许多生物学家则倾向于相信“绝对偶然性”假说。

3 “生命博奕”

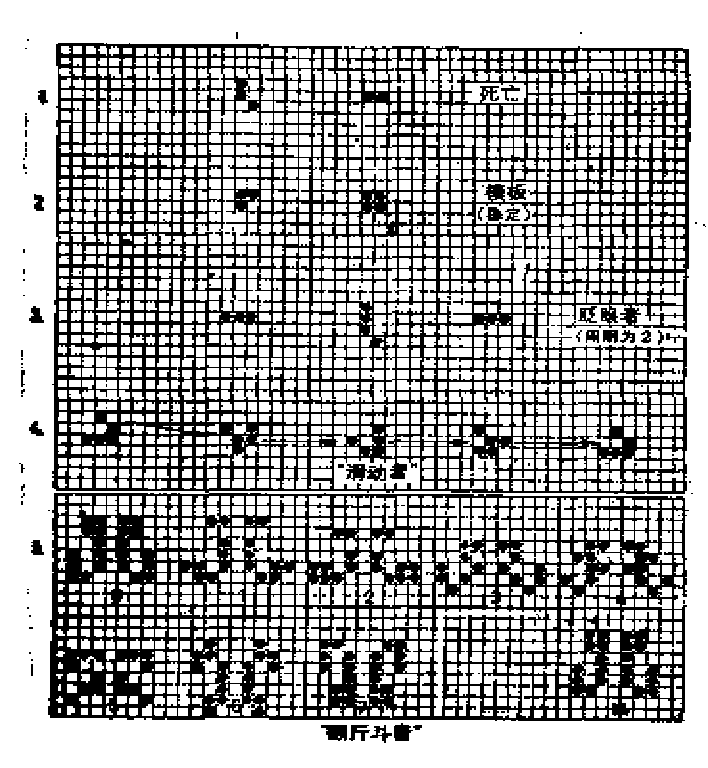
冯·诺伊曼关于自复制机的思想，已经激起了数学家对一种模拟增殖和生长的特殊类型的博奕发生了兴趣。乌拉姆在“论递归定义的几何对象与生长模式”一文中，描述了一些应用于细胞自动机的这种博奕；更新近一些，康韦引入了一种激动人心的博奕，他称之为“有生命的”，因为“它类似于某种生命有机体社会的兴衰和变化”。我在这里提到这些博奕，有两点理由：

(1) 它们极好地表明了，从非常简单的构造规则，如何能够获得难以想象的复杂性，这些复杂性是属于那些包含着许多功能特征的对象的。

(2) 它们恰恰类似于前述的三个关于生命起源假说中的第一个假说：由最初的构型预先注定了完全决定的进化途径，通过正确实施转移规则，它就由此沿这条途径前进。这些博奕一般具有某个有限的或无限的小格子，其中每个格子均被指定处于有限数目的给定状态中的一个状态，即“空态”或“占态”。另外，每个小格还具有有限数目的从而相邻小格影响着它自身的状态。把某种转移规则同时施于每一小格，则状态的模式以离散的时间步调发生变化。每次博奕的结果，不外乎是消亡、生长和静止的行为(即某种稳定的或振荡的模式)这三种情况之

一。表 3-1 展示了康韦的生命博奕的特征。模式可演变得极为复杂，以致除了非常少的简单起始构型以外，要预测任何最终结果都是极其困难的。尤其是，诸如“飞船”、“滑动炮”或“捕食者”(参见表3-1)此类具有“战略”重要性的结构，把某些特征添加到了这种博奕中，从而使得对其感兴趣的门外汉以及科学家都将为之激动，科学家也许会抽象出新的原理，以构造一种万能计算器，即图林机。

表 3-1 康韦的生命博奕



该博奕在一块"无限大"棋盘上进行。每个小格可采取"占"或"空"态。小格上可置上筹码以表明其"占"态。给定的小格状态取决于其相邻小格的状态。在康韦空间中，每个小格都有4个正邻小格和4个角邻小格，共计8个相邻小格。时间以离散的步子流逝，每一步都构成一次生育或一次"移动"。这些规则同时适用于所有小格。它们是：

1. 生存。与2个或3个筹码相邻的每个筹码都可以存活到下一代。

2. 死亡。有4个或更多个邻居的筹码从过于稠密的群体中死亡，即被移走。只有1个邻居或没有邻居的筹码，因孤立而死亡。

3. 出生。每个不多不少正好有3个邻居的空格子是生育格子。它可在下一次移走时放上一个筹码。

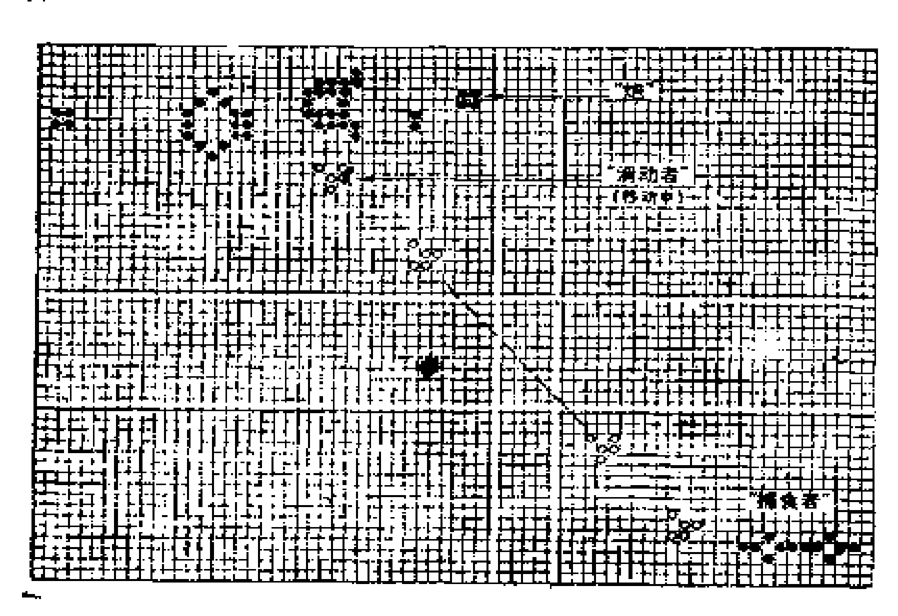
重要的是要理解出生和死亡是同时出现的。表中已表明了几种典型的构型。相连的走子表示为连接的步子。

三个一组，即死亡(1)，变得稳定(2)，或者振荡(3)。

滑动者(4)，即横跨棋盘的周期结构。(注意，其余的振荡子保持原位不动。)

振荡子(5)的周期为14。每移走7次，即上下颠倒一次。这个"翻斤斗者"是小利林斯发明的。

续上表



左上方的"滑动炮"发射滑动者，被右下方的捕食者"吃掉"。滑动炮是一个周期为30的平稳振荡子，每一周期发射一个滑动者。滑动者向右下方移动，被周期为15的振荡子即捕食者吃掉。

表 3-2 *tRNA*博奕

给定：对于每个博奕者，有一个用字母 *A*，*U*，*G*，*C* 表示的*N*(例如80)位四类字符的随机序列和一个四面体骰子，骰子的每一面代表一个字母。

目标：依次掷骰子，依据所掷骰子的结果把序列中相应的位置加以代换。每个博奕者都要尽量趋近具有尽可能多的AU或GC对的双链结构。一旦达到最先出现的

"完整"结构，博奕即行终止。每个博奕者都要记下配对数，其中GC对按AU对的两倍记数。在每一轮每个博奕者只能掷一次骰子，但他可以自由宣称自己的骰子掷得的某一个位置。

规则：(1)互补规则。两个互补字符(A和U或G和C)相邻时配对，即被联系在一起。

(2)协同规则。只对那些具有下列情形的互补对计数：即它们至少呈现为4AU或2AU+1GC，或2GC对的不间断序列。

(3)立体规则。由2维折叠可形成任一模式。折叠需使两个字符进入相邻位置，它将产生环。由于"立体"原因，环区的5个字符必然保持不配对。

进一步可能的约束是该链的端-端匹配条件。

对程序的说明：环太多的话，则使配对的可能性减少得太多，还导致对协同规则的限制。一个环——象在发夹构型中出现的那样——将把起始配对的机会限制在统计期望值之内。随机配对的较大涨落，即对期望值的正偏差，只可能在较短区域的特殊结合中发现。寻求这种区域，需要形成几个环。这种优势应归于存在一种很大的结合多重性，它与一个发夹构型相反对80个字符而言，最优结构就象三叶草。而且该博奕很好地模拟了图5所示的较长序列。

这个博奕的细节是由温克勒完成的，他的工作基于弗里斯科、艾伯茨和多蒂早期对结构的估计。

此博奕中所缺少的，是某些随机性因素。任何分子的复制过程，都会随机地出现"误差"。如果适当地选择，这些"误差"或"突变"可被看作进一步进化之源，即新信息之源。比起由一位"全知的创世主"已经置入起始模式中的信息来，康韦的博奕并不创造出更多的信息。该博奕不受任何其他影响，正确实施转移规则它就必然走向停止。

设计出作为"绝对偶然性"假说的例子的博奕，当然没有多

少价值。任何抽彩号或掷骰子都是合适的。如果引入适当的选择规则，这种博奕便可成为进化行为的代表。

考虑某字母序列，它可以代表一个核酸分子的碱基序列(见图3-3)：

AGUUCCGCAGGCU——→ν个字符

$\lambda=4$种

我们的任务是，在确定的选择规则指导下，由随机地变化单字符而获得某特定序列。"随机源"可用正四面体骰子模拟，此骰子的每一面分别代表A、U、G和C这4种字母之一。

"绝对偶然性"将相当于同时掷ν颗骰子，即每一颗骰子代表着ν个位置中的一个位置，而且同时相应排列字母。如果我们寻求某特定序列，那么平均需掷4^ν次才能做到这一点。例如，如果ν大约是80(即相当于最小的功能核酸分子——氨基酸连接物$tRNA$)，这场博奕将是"无止境的"(也是呆滞乏味的)，纵然如同抽彩号时不可能完全把"幸运号"的出现排除在外。

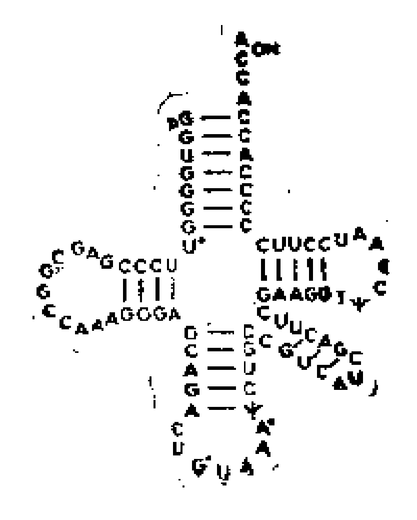
我们可以把更多的智力因素渗透进这场博奕中，只要我们引入进化标准并确定选择规则，通过它们迫使该系统与此标准相匹配。

我们所要选择的东西是什么？

信息如果不能保持到被读出，它就没有价值。任何属于热运动的非平衡结构都将瓦解。特别是携带信息的大分子序列正是(而且必须是)这种非平衡结构。这样一来，它们的"ur-语义学"仅仅能够涉及信息守恒，即生存。那些序列是有利的，它们提供进一步的信息以改进这种品质；增进复制速度或精确度、或者抵制分解的信息。任何自组织的选择机制，都必定对这些性质起作用。实际上，达尔文原理正是在进化的这个层次上起作用的。

极易被水解作用降解的核苷酸线性聚合链，由于生成由图3-3所示的、利用A—U和G—C的特殊互补作用的双链区域，从而得到有效的保护。这种力图取得最大的碱基配对（更倾向于包含较稳定的GC配对）的聚合物，因此将取得最好的生存机会。进一步的线索是，孤立的碱基对并不稳定，只有通过集结过程才能形成双链区域，这个集结过程包含了2GC或1GC与2AU或4AU配对之间的协同作用。而且，折叠时伴随着立体约束因素，从而限制了最大配对数，配对只有在某种特殊结构型式中才能形成。所有这些约束将在我们的博奕中通过特定的规则而加以考虑。

表3-2描述了这种进化博奕。其定量规则基于已知序列的孤立的低聚核苷酸实验数据。目标在于，在最短的可能时间中（即以最少的掷骰子次数），达到一种以最大的*AU*和GC配对数（其中更“有价值的”GC配对按较不稳定的AU配对的两倍计数）为特征的结构。获得的（二级）结构（如“三叶草”）型式相当重复地象博奕中的胜利者一样出现，并引人注目地与迄今由序列分析所核实的“自然”结构完全一致（参见图3-5）。大多数这种RNA分子都已经对其功能性质进行了选择（例如作为氨基酸连接物），而不是对其编码蛋白质结构的能力进行了选择。对这些分



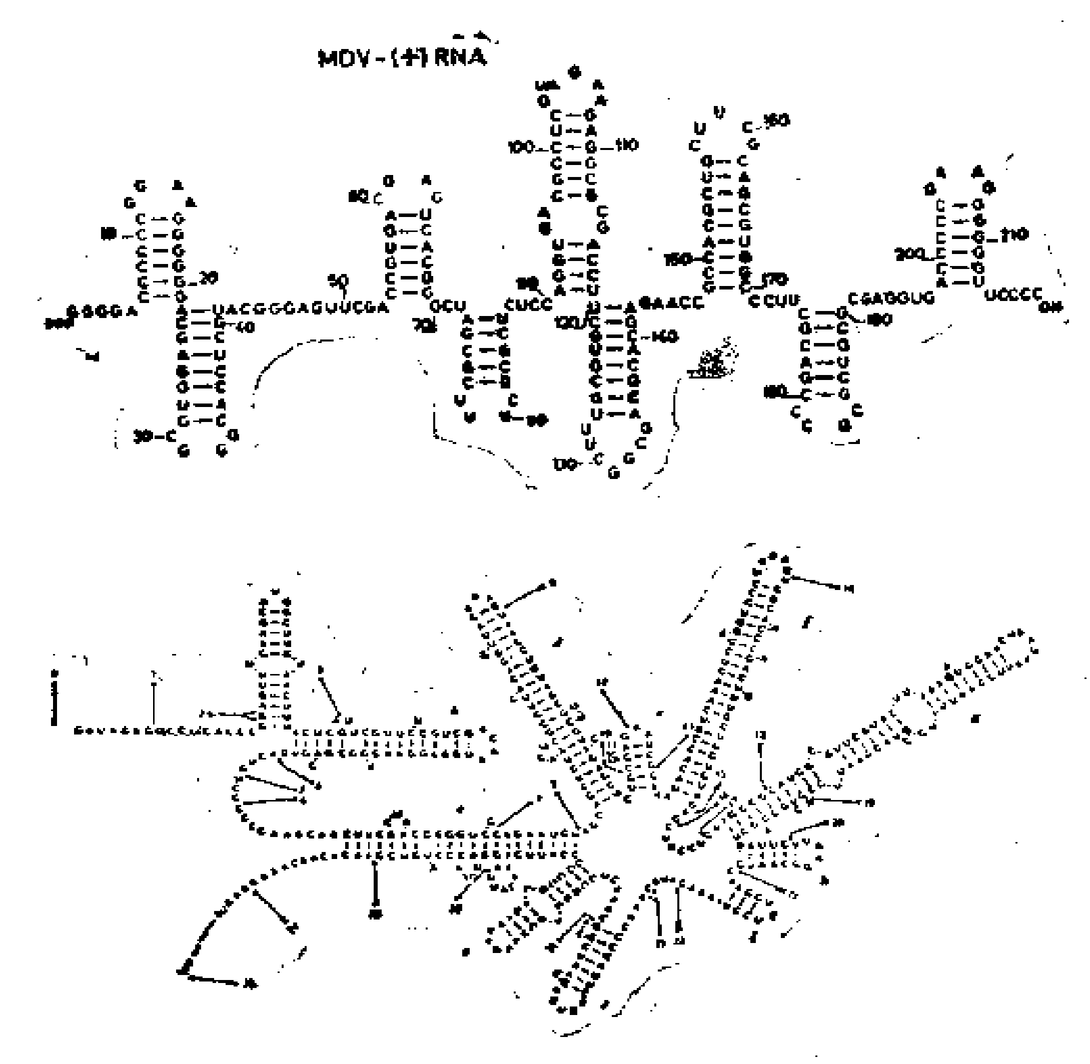


图 3–4　功能 RNA 分子的"二级"结构

(*a*) 转移 RNA，即丙氨酸的连接物。

(*b*) 运用 $Q\beta$ 噬菌体的特定复制酶在进化实验中获得的"中变体"。

(*c*) 由噬菌体 MS2-RNA 的核酸酶消化获得的"外壳"基因。

这个基因编码 129 个氨基酸的序列。

所有的结构都在空间进一步折叠。

子来说，稳定性的确具有本质的进化价值。这一点能够在试管试验中直接验证，在试验中某个功能性质是属于特定的选择株系的(参阅第5章)。

表3-2中描述的博奕，几乎模拟了自然选择过程中的极端现象。如果选择优势是指正确地占据了任何单个位置，而与其他位置的状态无关，那么就达到了一种真正的极端情况。在没有诸如协同性或空间阻碍等任何附加约束时，配对就会类似于这种情况。为了占据指定的ν个位置，我们仅仅需要次数最小的$\nu(\lambda-1)$次试验。(对ν个位置中的每一个位置平均尝试λ次，但是人们可以从某起始模式入手，此时可预期有ν/λ个位置可以被正确地占据。如果"配对"是唯一的约束，那么ν仅仅包括一半可配对位置，这正好与另一半匹配。)要领会$\nu(\lambda-1)$这个数，须与λ^{ν}这个数比较，λ^{ν}是在没有任何选择规则时无条件的尝试次数(例如，对于$\lambda=80$，$\nu=80$，则大约为240比10^{48})。这种非协同选择过程的极端情况难得在自然界发现。它将显得是决定论的，因为它必然达到与该途径无关的某个真正的最大值。主要在表现型层次上很明显的协同性要求(例如，在蛋白质的三维结构中以及它们在控制功能中的相互影响)，将大大地增加试验次数。它将使得途径出现许许多多分支，这就使得最后结果将依赖于不确定随机事件的历史顺序。这种最优化程序明显地不同于热力学平衡过程，热力学平衡中的真正极值(自由能取极小值)是独立于途径而达到的，而且仅仅是状态性质的特征。

如果我们现在返回到"自然选择"事件，我们就必须找出，在此博奕中掷骰子并按照规则进行选择的博奕者，是如何被分子机构中内在的机制所代替。如果要把达尔文原理推广到分子水平，我们就必须把它与分子性质联系起来，并从更基本的物理学原理中将它推导出来。

4 达尔文和分子

4.1 稳定性和选择

选择性的自组织能力必须与物质的特定动力学性质相联系，并且只有在特定环境中才能发现。

让我们考虑某化学实体i，i可以由某种化合物、某个大分子序列、或者甚至某个具有协同反应行为的相互作用物种的系综所组成。这种实体平稳地形成或分解，它们可能是一个复杂的多分子过程，并包含一些平行的或串联的反应步骤，这些反应可以是任意阶的：

$$\Longrightarrow i \longrightarrow 。$$

i的群体变量记作c_i，其变化δc_i记作x_i。如果相应的产生和消失(包括分解或运输)的净速率正好相等，则浓度c_i取某种定常值。特别是，如果两个箭头连接着完全相同的状态，即如果一个状态是另一个的逆状态，则c_i的定常性对应于一种平衡分布。

我们把注意力集中在i的涨落上，涨落出现在群体变量中并影响着净生成速率和净分解速率。如果速率的变化与群体变

量有相同的符号，我们可把这种情况记作“+”，如果符号相反时记作“-”，又如果这个速率与x_i无关则记作“0”。对于耦合反应，为了决定速率项的涨落响应的符号，需要作完整的系统分析。9种可能的情况可以用“支付”矩阵的形式表示：

		生成		
		-	0	+
分解	+	S_{11}	S_{12}	D_{13}
	0	S_{21}	D_{22}	I_{23}
	-	D_{31}	I_{32}	I_{33}

矩阵的主对角线上包含了三个最基本的结果：稳定态(S)、随机漂移(D)和不稳定性(I)。其中的下标有助于澄清不同的情况。

任何接近真平衡的过程都将落入S_{11}类型中；下面将联系埃伦菲斯特的缸子模型对它详加讨论。对开放的非平衡系统，第一行和第二列代表了比较普遍的反应行为：一级分解过程以及产物对反应物无任何反馈。其他两行和两列分别要求更专一的性质。尤其是在生化系统中能够发现它们，它们可以包括自动、交叉或循环催化(第三列)，产物抑制(第一列)，催化部位的底物饱和(第二行)或诸如变构调节所展示的更复杂的控制（第三行和列)。最后两种情况仅仅在有限的底物浓度范围内才普遍有效。

“支付矩阵”反映出稳定状况和不稳定状况的一种对称性，即S_{11}、S_{12}、S_{21}与I_{33}、I_{32}、I_{23}比较。按照李雅普诺夫标准，这6个结果类似于“易判定的”情况。如果在某稳态($dc_i/dt=0$)的邻域中，我们用$x_1, \ldots, x_n$的幂级数展开这个速率方程系统，便可

写成如下形式：

$$\frac{dx_s}{dt}=\sum_{i=1}^{n}p_{si}x_i+F_s(x_1,\ldots,x_n)。 \quad (1)$$

函数F_s包括了所有非线性(而没有任何线性)项，因而能够对充分小的x_i收敛。设λ_i是该线性系统的本征值，即是如下方程的根

$$|P-\lambda E|=0;\quad P=\|p_{ik}\|;\quad E=单位矩阵。 \quad (2)$$

于是，不论F_s的形式如何，立即可以区别两种情形，即这些通解($x_s=0$)是"渐近稳定的"，还是"不稳定的"。业已表明，对这两种情况，系统(1)具有当$t>0$(稳定)或$t<0$(不稳定)时均可收敛的级数(包括在时间t指数与多项式的乘积)形式的解族。如果所有$Re\ \lambda_i<0$，则结果稳定；如果其中至少存在一个λ_j，使得$Re\ \lambda_j>0$，则在λ_i中出现不稳定。如果不存在正的λ_i，则可能不那么容易作出判断，但至少存在一个λ_k，且$Re\lambda_k=0$(即一个为零的根或一个虚根)只有在"支付矩阵"的反对角线上($D_{13}D_{22}D_{31}$)才可能包含这些"可疑"情况，尽管"可疑"情况并不唯一地由它们代表。若按照李雅普诺夫方法检查这个展开的方程可能会揭示出稳定的情形或不稳定的情形，这依初始速率方程的特殊形式而定，尤其是其中的线性项。普里戈金在其演讲中已经给出了一个关于稳定性标准的非常一般的公式，这个表达式应用于远离平衡的稳态。① 他把涨落效应归于耗散项

$$\delta_x\sigma=\sum_k\delta J_k\delta X_k,$$

式中σ是来自系统内不可逆过程的熵产生dS_i/dt，J_k是广义流，

① 注意他定义的局域平衡，它允许温度、压力、以及熵、化学势等变量具有如同在平衡热力学中那样的意义。这个假设应用于本文中的任何远离平衡的化学过程，均保持有效。

X_k 取其共轭的广义。对于化学反应系统，流和力由(非向量的)反应速率 V_k 及相应的亲和力 A_k 来表示。从化学势中导出的亲和力，包含着对数的浓度项，导致了对 i 分解的 X_i 的正响应，以及对 i 生成的 X_i 的负响应。(注意，对于处在非平衡定态的化合物 i，至少有两种亲和力分别表示 i 的生成和分解。)如果我们现在分析 $\delta V_k \delta A_k$ 分别对这两种过程的贡献，并与前面引入的"支付矩阵"进行比较，我们发现同格兰斯多夫—普里戈金的标准完全一致：对所有 S–情况，$\delta_X\sigma>0$；对所有 I–情况，$\delta_X\sigma<0$；以及对于 D_{22} 有 $\delta_X\sigma=0$。对 D_{13} 和 D_{31}，生成和分解的贡献具有相反的符号，所以这三种结果均是可能的，这取决于速率项的特殊形式及大小。

4.2 随机模型

无论是采取一种稳定的分布，或者随机地在不同状态之间扫掠，还是作为一种不稳定性的结果发生不可逆变化，多样性都将成为进化的自组织的最重要前提。对这些行为原型进行比较，并对它们的随机本性展开讨论，会进一步说明并突出这一点。

稳定行为最好以埃伦菲斯特缸子模型来说明，一些物理学家也把这个模型称作狗—跳蚤模型。我可以扼要描述这个经典的统计力学难题，鉴于要进一步讨论其他模型，使我不能完全省略掉它。考虑两口缸子，或两个匣子，在其中任意分派 $2N$ 个球，每个球分别标以数字1，……，$2N$ 中的一个数字。另外，有一个匣子装有签条，每个数字有且仅有一根签条。为了保持所有球有相等的先验概率，在博奕期间必须保持这组签条的完整。博奕的程序是进行抽签，并把抽得签条所代表的球从一个匣子

转移到另一个匣子。众所周知的结果是球等分于两匣子中，且与起始分布无关。在每一个匣子，都将平均有N个球，其平均涨落 $n=\pm\sqrt{N}$。

该博奕模拟两种相等自由能状态之间的(化学)平衡，当然容易把它一般化，并推广到任意K种状态之间的分布，这K种状态有不同的被移居权。重要之处在于分布的稳定性。其涨落n出现的概率P_n为

$$P_n=2^{-2N}\frac{(2N)!}{(N+n)!\,(N-n)!}\approx\frac{e^{-(n/\sqrt{N})^2}}{\sqrt{(\pi N)}}.$$

特别有，$P_{n=0}=(\pi N)^{-1/2}$，而 $P_{n=\pm N}=2^{-2N}$，它也是任一单微观态的概率。对大数N(这对大多数统计力学难题是现实的)，这种概率是如此之小以致巨涨落(几乎)从不出现，而"小"涨落(即在$\pm\sqrt{N}$的平均极限之中)则非常频繁。这种行为是由涨落的自调节性质带来的。它们越是大，则它们的逆转概率就越高。

例如，这种性质在代表D_{22}情况的抛硬币博奕中被遗漏了。此处，先前发生的涨落对下一个结果的概率毫无影响。如果在这种博奕中，我在一连串的成功之后开始计数，并且将这些成功事件也计数在内，我就会以某种不公平方式偏爱我的位置，因为对于任何进一步的抛掷，我仍然有成功一半的概率。在埃伦菲斯特博奕中情况不一样，涨落漂移不能自我控制，所以分布随着时间推移而平稳地变宽。

与埃伦菲斯特模型相联系的涨落行为，正是达尔文选择机制的"逆指示"。从第3章中描述的巨大复杂性方面来看，它绝对不许运用稳定的平衡分布。这样一个系统不可能作出独特的选择，因为依据其相对自由能含量它会占据所有的可选择状态。所需要的反而是包含不稳定性的机制，在生长限度的约束下允

许无优势的拷贝完全消失。

这种行为可以用另一种博奕来模拟。准备N^2个球，其中数字1至N中的每一个有N个拷贝，从准备的N^2个球中对每一种数字取出一个代表即总计有N个非简并拷贝，并把它们置入一匣子中。博奕中的程序是，任意从匣子中取球，并轮流地置它们于清除或复制两种命运中。这些命运的严格轮流，保证了在整个博奕期间，匣子中有恒定的N个球的群体。必然的结果将是：选择了一个显示出N重简并的数字。

实际上，这场博奕可用两种不同方式进行。可以随机地而不严格地轮流清除和复制，那么，匣子中球的数目将不会保持恒定而围绕平均值N发生涨落。这些涨落不象在埃伦菲斯特模型中那样，它们不能自我控制。这样一来，随着时间推移，球的数目将不断地离开N(朝两个方向)，直至某"涨落突变"灭绝了这整个群体。这种命运是不可避免的，就象在第一种情况中命运是唯一的选择一样。在涨落突变发生之前，起始的N个非简并数字将下降到几个(或者甚至一个)，但这几个数字将会产生许多(对整体N的平均)拷贝。

这种博奕可以看作为一平稳的马尔科夫过程。由杜布引入并由巴塞洛梅运用于类似的群体遗传学问题上的一种随机处理方法，定量地揭示了所描述的结果。对球进行N^2次的清除或复制后，某个涨落突变出现的概率达到e^{-1}。这种涨落突变类似于从一个球扩展到N个拷贝，即在恒定群体中的唯一选择。假定平均复制时间等于球的平均寿命τ，等价于此博奕进行了N次试验。因为N个球体中的每一个，在时间τ内均具有一次复制或者清除的机会。

在这种博奕中，各个球的性质不可区分，因之无法预见哪一个球将被选择。我们仅仅能说，存在一个数，且对恒定的N

是仅仅存在一个数，它将保留下来。这种性质是该种系统的内在性质，归因于该种特定类型的涨落行为。这些涨落漂移着，但是由不可逆清除的奇点引进一种不对称性，这种不对称性倾向于保留高度简并的拷贝，奇点出现于群体数为零处。它具有一种与涨落的“真正的”放大作用相类似的效应，这将要求复制速率超过清除速率，因此它不可能在上述的特定博奕中出现。

这种形式的博奕模拟了达尔文选择吗？

达尔文为他的自然选择原理采取了斯宾塞的“适者生存”的说法。有争议的是，如果最适者只不过是指生存的事实，即如果除了由生存的事实而无其他方式去确定“适应”，那么，这种形式的自然选择原理显然表示了一种同义反复，即是说“生存者生存”。

对于我所描述的博奕，这的确是真的。起始的不决定性不是由于知识的缺乏，而是由这样的事实决定的，即所有竞争者的清除和复制的先验概率都相等。

把这种假设应用到分子的分解和复制过程，不仅是极不现实的，甚至还会在恒定群体的约束下把选择排除在外。所有的分子过程都有热“噪声”。于是为了正确(没有错误)地复制任何字符序列，便会有如下要求。或是要求相互作用的能量很大(直到不现实的程度)，以精确地甄别字符；或者——假定已知相互作用能相对地小——没有热“噪声”，字符序列代表着一条信息。事实上，总是只有一部分复制序列是正确的，即$Q<1$，而与进化程度无关。在进化期间单字符识别的误差率减小到小于10^{-10}。

我们可以解释不完善的复制，并相应修改我们的博奕。从匣子中取出的任一第R个球，在复制阶段可补充以某个具有新数字的“新”球，这个新数字$>N$(例如，若$Q=0.9$，则$R=10$)。如果离去和进入该博奕的球的总速率是相匹配的，那么对于任

何存在于该匣子中的球来说，清除的概率将超过复制的概率。这是显然的，因为新数字稳步地加入进来。因此，任何数目一定的群体都只可能具有有限的寿命，这意味着稳定选择绝不可能，至少对所有球的清除和复制的先验概率是退化的情况下是这样。如果各个种类的反应行为不同，才可能在有噪声的情况下进行可重复的选择。于是“最适应的”实际上指的是某种个体性质，这一点现在可以由独立实验加以核实。它涉及到再复制速率、精确度、以及寿命的某种最优结合。带来那些最适物种的涨落，将真正被放大，而那些较不优越的拷贝将衰退。这结果是某种由个体行为引起的具有强烈倾向的选择过程。第3章讨论的*tRNA*博奕提供了一个富有启发的例子。“博奕者”现在已由该系统所固有的自组织机制所代替。

4.3 唯象表示

在分子水平上的选择和进化，本质上都是随机过程。它们通常开始于某一个拷贝①，这个拷贝是在一个不明确的基本过程中产生的。如果某个优势拷贝出现了，它必须在确定的选择成为现实之前达到一定的丰度，这取决于它的选择优势的大小。从另一方面来说，这种不确定性在本质上在于拷贝选择，而不在于放大过程。

选择的机制可以用决定论的理论来描述，这理论本质上以三个唯象参数为基础：

A：速率参数标征模板指令的复制过程。这种过程具有内在的

① 通过不确定的量子现象的放大而导致的宏观不确定性的可能性，很早以前就已被约尔丹考虑过。

自催化本性，因此导致拷贝的增加。

Q_i　品质因子描述复制的精确度。于是精确复制由 A_iQ_i 项标征，而$A_i(1-Q_i)$描述错误拷贝 k 的产量，k 与 i 有关（属于不同类的"亲和度"）。拷贝 i 对任一拷贝 k的产量的贡献，可以用概率分布 φ_{ik} 来描述，且有

$$\sum_{k\neq i}\varphi_{ik}=1。$$

D_i　也是一个对拷贝i分解的速率参数，这些参数代表了特定的速率。因此，在唯象方程式中出现的速率项，是 A_i、A_iQ_i 或 Di 与群体变量 c_i 的乘积。所考虑的反应类型属于前面引入的类型 D_{13}，否则这三个参数可以是 c_i 的任意函数。特别是，量 A_i 总是包含高能(单体的)反应物(它们可以或者不可以加以缓冲至恒定水平)的浓度的某种化学计量函数。

在速率方程中，三个参数作为一个复合项出现

$$W_i=A_iQ_i-D_i$$

可以称它为选择价值。本质上正是这个量在自然选择的竞争中是决定性的。

选择价值展示了阈值性质($A_iQ_i\gtrless Di$)，但是，作为与纯粹的分离相区别的真正的选择，需要某种平稳的可调阈值。这一点可以通过引入控制而办到，即引入恒流——限制物质和自由能供应，或引入恒定的总力，即缓冲反应物和反应产物的总浓度(恒总亲和力)。

在自然界中绝不出现遵循这些明确约束的选择过程，正如蒸汽机不可能在理想化的卡诺循环条件下工作，然而可用卡诺循环决定它的最大效率。

在这种博奕中，由复制和清除的严格轮换保证了恒定群体。在化学系统中，为平稳地抵消过量产生 $\sum_{k=1}^{N}(A_k-D_k)x_k$，我们必须调整流量。展示了选择行为的最简单系统，是那些选择价

值 W_i 为常量的系统①。在恒定群体中，所有其选择价值超过了平均生产力 $\bar{E}=\sum_{k=1}^{N}(A_k-D_k)x_k/\sum_{k=1}^{N}x_k$ 的物种，都将生长，而所有其他的物种都将衰退。从而，它们将 $\bar{E}$ 移向由系综极大值 W_i 决定的值：

$$\bar{E}\longrightarrow W_{max}$$

这种"被选择的"物种将支配该群体。伴随它的是某种重复出现的突变体的整个家族。突变体离被选择的拷贝的"亲和度"的距离越大，出现突变体的速率就越低。如果在非成群的、更为疏远的亲系中，某随机的有效涨落最终带来了优势拷贝 $(W_i>\bar{E})$，那么它将生长，并导致打破先前的(亚)稳定群体。这种系统中的进化由单调序列标征：

$$W_{1m}<W_{2m}<\cdots<W_{最优}，$$

可以表明，在恒流时，某优势突变体的出现相应于耗散项中的某个负涨落，即

$$\delta_X\sigma<0。$$

在恒反应流时形成的重新组织，将产生出较高程度的有序，这反映在内部熵的减少上(如果我们假定在自由能量方面，这些竞争者是简并的)。当然，内部熵产生总是保持为正，只有消耗外部流量才可能使内部熵稳步减少。

一般在线性系统中会遇到这种简单的单调性极端行为。② 在耦合(仍为线性的)反应系统中，同样的标准适用于由群体变量仿射变换获得的那些简正模式。循环催化由属于循环上的 R

---

① 那些系统必须总是远离平衡的。接近平衡时，所有速率方程(与其形式无关)可以线性化，并产生恒为实数的负本征值(这是表达了微观可逆性的翁萨格关系的结果)。

② "线性"项指的是 W_i 项。非线性项只有通过外部控制才能引入，从而它在 E 项中与负号一起出现。

个简正模式中的一个正的本征值来描述。其余的$R-1$个本征值是负的，它们描述循环系综的有关群体水平之中的某种平衡。

还有一些其他类型的重要的非线性反应系统，该简单的单调优化原理仍然适用。这样的系统通常包括若干个物种的系综，它们通过非线性反应耦合联系起来，在循环联接聚集之后，导致总系综整体上表现出象某一个体的行为。第5章中更详细地描述了一个有关转译起源的模型系统。

从更一般的观点来看，最优化不必包含单调的时间行为，"适者"也不必是指相应群体变量的系综极大值。

这些唯象参数(无论它们的特殊形式如何)描述选择和进化行为，如果它们描述的系统是由如下特征所标志的：

(1) 代谢，至少由两个独立的、具有正亲和力的速率项来表示。它们涉及到在高能材料连续向低能材料转变中的生成中间物。

(2) 自我复制，是这些中间物的内在性质。这种品质由特殊形式的放大速率项(需 $A_iQ_i>D_i$)来表示。

(3) 中间物的突变性，由品质因子 $Q<1$来表示。这种条件需要存在可选择的中间物，其中的一些中间物必须是自我复制的。

这些前提对于选择是必要的，但（正如将要表明的）不是充分的，至少就向已知的细胞生命水平的无止境的进化而论是这样。

在这种联系中具有特殊重要性的是品质因子Q。选择的阈值性质需要Q的极小值，这取决于该速率项的相对方差，即

$$[Q_m]_{min}=\frac{\bar{A}_{K\neq m}+D_m-\bar{D}_{k\neq m}}{A_m}$$

式中下标m指的是已选择的物种，并对全体竞争者m求平均值。

另一方面，Q 涉及到代表了ν个字符的一条讯息的某个信息载体。我们给每一字符类赋以一个基本的精确度因子 q_i，q_i 可以用字符丰度和相互作用的自由能来表达（参见第 5 章）。对于 λ 种可区别的类别（如果识别以协同作用为基础，则很可能包括配对结合），我们得到

$$Q_i=\prod_{k=1}^{\lambda} q_k \nu^{ik} \quad \sum_{k=1}^{\lambda} \nu_{ik}=\nu_i$$

或者在最简单的可能情况中，有一致的（或平均的）精确度q：$Q_i=q\nu^i$。这种关系包括能够安全复制的字符上限ν_{max}，给定某一确定的速率参数的方差，假定对任一字符k有$(1-q_k)\ll 1$：

$$\nu_{max}=\frac{|\ln Q_{min}|}{1-\bar{q}}; \quad \bar{q}=\frac{\sum_k \nu_k q_k}{\sum_k \nu_k}; \quad \sum_k \nu_k=\nu。$$

除了 Q_{min} 非常接近1（即对每一非常小的选择优势），$|\ln Q_{min}|$项数量级为 1（否则它将更小）。

换言之，速率项有比较大的方差，选择不可能明显地影响ν_{max}。即使被选择的物种的速率参数与其竞争者的速率参数相差几个数量级，由于是对数关系，它对ν_{max}的效应将是无关紧要的。另一方面，对 $\bar{q}$ 的任何改进，都会极大地影响ν_{max}。在进化期间，遗传信息复制的单字符品质因子，即核苷酸识别的 $\bar{q}$ 值，从非酶协同的核苷酸相互作用的大约0.99，增加至体外的酶识别的$(1-10^{-5})$，在较高级的有机体内超过$(1-10^{-10})$（包括"修复酶"的密码核实，以及由多基因表达带来的冗余）。遗传信息的长度能够相应地从大约 100 个核苷酸增至在较高级有机体中大大超过10万亿个。主要在起始阶段进化的优化与速率有关，速率用W_i或 $\bar{E}$ 的绝对值表达，在这个阶段必须适应酶的机构。后面的阶段，更多地由信息容量（ν_{max}）的增加所支配。相应地要求

提高复制的精确性。这使得物种变得越来越不依赖于它们起源的特定环境的先决条件。与平均产生相联系的选择价值，仅仅反映了对特定环境(包括外部因素以及潜在的内部耦合)的"适应"。

4.4 进化实验

选择是某些明确定义的物质条件所固有的，这个事实使得达尔文原理能够从更基本的物理学定律推导出来。于是，该原理并不只是反映了特定的进化历史事件，应用它时也不需要先存在任何形式的"生命"(正如通常所认为的)。给定适当的边界条件，选择和进化看来是一种不可避免的过程。所以，在试管中再现这种过程是可能的，至少就原则步骤而论是这样。拷贝选择的非决定性和优化过程与途径的依赖关系，可能会排除任何历史时期中的过于相似的复制。

一个典型的类似于达尔文选择的试管实验，已由施皮格尔曼及其同事实现。这个实验如图3-6所示。流量控制使得高能反应物(即核苷三磷酸*ATP*、*UTP*、*GTP*和*CTP*，或者活化的氨基酸，如果涉及到蛋白质合成的话)调节在恒定水平，以及把聚合反应产物调节在某个恒定的总群体数。利用半透性的过滤器，溶剂、单体反应物和聚合产物的供给和移去是分开进行的。这种形式的实验几乎可以由选择博奕准确模拟，在选择博奕中假定先验概率的分布是不均匀的，并轮流清除和嘈杂地复制。

对分子自组织的系统实验研究，仍处于起步阶段。迄今只涉及到，在核酸层次上的进化过程，其中以酶作为环境因子。这种巨大的结构多样性及其在表现型表达中的作用，正如包括噬菌体和人造核糖核酸的研究中所揭示的，是颇令人意外的。这

些结果，大大修改了我们关于蛋白质和核酸的独特结构之中识别的相互关系的观点。它们强调了，在一个选定的 *RNA* 分子中，要求碱基配对区域中有一定的对称性（参见图3-5）。这种对称性允许同一个酶能够识别二条互补链。这些发现强烈地倾向于蛋白质-核酸相互作用的某些非线性模型。

那些实验，还必须以对酶功能进化的研究来补充。要实现包括蛋白质合成的实验更为困难。连续的多肽合成需要可靠的（若可能最好是自动化的）体外方法。进化程序可以主要靠计算机去执行。它只需要不时地、但是系统地检查由代表基因型层次的程序所带来的随机变化。这种策略可能并不试图从头合成

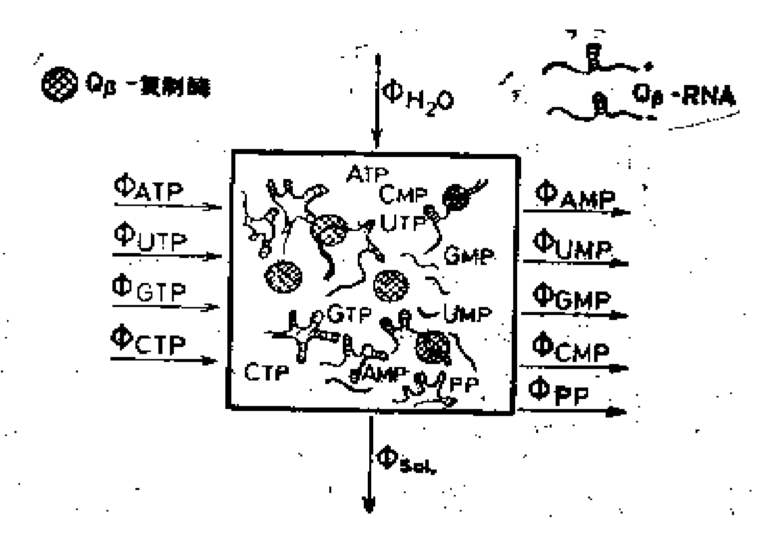


图 3-5　4.2节描述的选择博奕模拟一个进化实验，这个实验在原则上最先是由施皮格尔曼运用 *Qβ* 噬菌体 RNA 和复制酶实现的。

"匣子"中包含了反应混合物，为模板指令的RNA合成提供最适条件（pH值，离子强度，等等）。把恒定水平的复制酶即该种噬菌体的专一的*Qβ*复制酶供给这个系统。全部流量是被控制着的：诸如四种核苷三磷酸高能反应物的流入量，溶剂的流入量，以及低能分解产物的流出量和溶液（其中包括聚合反应产物）的流出量。引物 *Qβ-RNA* 作为模板，引起聚合反应。实验可在恒定流量或者恒定浓度水平（即恒定的总亲和力）下进行。在施皮格尔曼的实验中，通过"连续转移"接近后一个条件。

“整个”生命体，而宁可成功地代替单一的一些组分，这些组分属于另外一个完全的生物合成系综——这类似于“客串”唱片的原理，其中只有一件乐器是要由一个业余爱好者来演奏的，而演奏组的其余部分全是由职业演奏家组成。设计那些实验并非用来在试管中产生“矮人”，而是用作检验作为各种自组织进化系统模型的基础假设。

5 实在和模型

一个理论要么是正确的，要么是错误的，二者必择其一。而一个模型则有第3种可能性，即它可以是正确的，但却是不相关的。

现有的理论解释进化原理以及它们的物理学定律的基础。它包括自组织的起源及其独特的结果——“生命”。而且，它还为评价实验室实验提供了一个定量基础，从而可能会指导我们通过排除不适宜的选择而得到恰当的模型。然而，这种理论将决不会再现进化的精确历史途径，“决不”是可能选择的巨大多重性的结果。

模型必须满足两个要求：

(1)它们必须从实在前提出发。

(2)它们或者它们的结果，必须与现在已知的事实一致。

这些前提主要是化学性质的。在前生物阶段产生的多种多样的化合物，其中包括了生物大分子的全部最根本的前体，特别是自指令的、催化的结构，这些结构或多或少地与随机性的聚核苷酸及多肽有联系。这可以由模拟了适当条件的实验室实验加以演示。至今并不是所有涉及到这个进化阶段的问题都得到了解答。不过，这些问题本性上是典型的化学问题，在原则

上量子力学能够完全解释作为生物复杂性基础的物质多样化。我们将讨论自组织化学网络的模型，先假定存在核酸和蛋白质类型的聚合结构，但此种结构是未组织的，在很大程度上是不可复制的。与实际存在的事物有关的一个重要问题是其丰度或浓度。

海洋的的确确曾经能供应过浩瀚的"汤"，但是，正是由于"供应"这一缘故，生命很不象从大多数的海洋中开始。合成大分子需要高能原料达到某个浓度水平。以单体反应物去饱和海洋，对利用能量资源将是极不经济的。而且，自组织是一种涉及不同特定作用物之间相互作用的协同过程，因此依赖于某种"集结作用"。这就注定了是滩涂、池塘或者甚至泥潭，而不是海洋，可使任一系综保持在一起并保护它不受稀释。而且正如福克斯和其他人的工作已经表明的，诸如温度的大涨落、表面催化等等，它们能为大分子合成提供更有利的条件。事实上，那些条件正是在实验室产生出功能性的高分子量的类蛋白化合物的条件。

5.1 真实自复制

第4章所描述的选择博奕，代表了自组织的达尔文系统的最简单模型。我们可以对每一"i"类球赋予一个选择价值W_i，但在那些球之间没有任何进一步的耦合作用，从而W_i与任何群体变量无关。整个群体的不变性受到外部生长控制的影响。

在恒定培养基(这需要调节环境因子和总的群体密度)中微生物的无性增殖，可由此种模型来描述。

我们在分子水平上发现了任何反例了吗？

引用适合这种自我复制机制的化学结构可能是容易的。然

而由于明显的原因，自然界并没有利用它们。大分子序列的进化势必以它们编码信息的能力为基础。这至少需要两类字符，即两类单体反应物。如果这两(或更多)类字符的自然丰度是这样的话，即两者都以可比的先验概率参与到聚合序列中，那么它就会对应于某个非常独特的意外事件。否则，最频繁出现的字符的颇为一致的序列便会具有最高的选择价值，它便会超过任何其他非一致序列而生长起来。

事实上，我们可预期，四种核苷酸的最初表现是极为不同的。腺嘌呤(A)在还原性大气中比在含氧的类似条件中的形成要容易得多。正如最近表明的，甚至通过直接的HCN缩聚就可以获得它。存在含氧化合物时，这个丰度可以变动。正如最简单的羟基嘌呤代表物6-羟基嘌呤，次黄嘌呤，即次黄嘌呤核苷(肌苷Ⅰ)的碱基组分，必定已经出现。肌苷在与胞嘧啶(C)的互补作用方面类似于鸟嘌呤(G)。完全可能的是，只是在化学进化的后期阶段，"I"才被"G"取代了，因为GC配对比IC配对稳定。而在早期阶段，"I"也许更受偏爱一些，因为它在与其他碱基相互作用时，类似于某种"伏笔"。在某些转移核糖核酸中的反密码子区域，仍在利用这种性能。丰度的宽分布有两点理由：(1)嘌呤(A，I，G)和嘧啶(U，T，C)的稳定性的差异；(2)通过聚合链中最近的近邻相互作用的个体稳定化效应。这种核苷酸的"堆积"是碱基配对协同作用的主要原因。所有这些影响，会强烈倾向于在某线性自我复制机制中出现一致序列。

5.2 互补指令

保证建立起某种巨大的大分子信息贮存能力的最简单诀窍是利用互补性，而不是利用自我指令相互作用。除非丰度高的

字符引入了一种相应量的频率较低的互补字符，否则丰度高的字符不可能复制。以这种方式，该系统至少能够发展起二元字母表。

以互补相互作用为基础，核酸能够进化到什么程度？

一个利用互补指令的交叉催化效应的核酸自组织复制的模型，已经从以第4章所描述的理论为基础发展起来。它适用于非酶模板指令，以及酶催化的复制，因此，酶可被看作某个环境因子，而不是某种进化的反应产物（参见图3-6中描述的$Q\beta$—实验）。例如在恒定总群体，选择动力学由耦合的微分方程系统来描述。每一物种，由一对涉及到两条互补链的方程表示。由恒定总群体的约束所引入的非线性项总是负的。应用于某种线性自我复制物种的选择价值W_i（$W_i=A_i\cdot Q_i-D$），现在被如下矩阵的两个本征值所代替：

$$\begin{pmatrix} -D_{+i} & A_{+i}\ Q_{+i} \\ A_{-i}Q_{-i} & -D_{-i} \end{pmatrix}$$

我们把本征值记作$W_{\pm i}$，其中的一个本征值可以为正，即如果$A_{+i}Q_{+i}\cdot A_{-i}Q_{-i}>D_{+i}D_{-i}$描述互补对的生长方式的话。另一个本征值总是负的，它描述了一种“平衡”方式，即正链和负链的群体变量趋近某个固定比值。最后的结果又是选择一个物种，该物种包括补体及其突变体家族。这一物种由某个最大选择价值加以区分，即

$$W_{\pm m}>\bar{E}_{k\neq m}$$

式中$\bar{E}_{k\neq m}$是超额产生的残基系综的平均值，下标k指的是出现在选择过程期间的任何时间间隔中的除了m以外的所有物种。它最终适用于随着群体数大于零而出现的所有突变体m。

对于我们的“生物信息的起源”问题，两个结论具有特殊重要性。

(1)选择价值现在指的是这样一项，它包含速率参数 AQ 和 D 的等比中项。所以，它可能以相同的权去选择两链中任一链上出现的优势。如果是速率或时间常数的等差中项而不是等比中项具有决定性作用，就不可能这样对两链同时进行优化作用。速率参数无论在何处受到结构性质（诸如 $tRNA$ 中的稳定性，或者诸如 $Q\beta$-RNA 中由复制酶的特定识别）的影响，对等比中项最优化的要求会带来两链的某种对称性，正如已由实验事实再度核实的(参见图 3-5 中的结构)。

(2)由纯粹的竞争线性选择过程形成的信息容量，被限定为一个单物种(包括两链)的字符含量。只有不断出现罕见的这种特定物种的突变体，才可能继续进化进步。

可以靠这种方式起源的最大信息容量，对应于量 ν_{max}(参见第4章)，从而严格依赖于单字符品质因子 q_i 的平均值。对这个简单的互补指令模型，q 很容易与可测的相互作用参数联系起来，这种参数诸如在任何字符 i 和 k 之间的复合形成的自由能 ΔG_{ik}，或者稳定常数 $K_{ik}=exp(-\Delta G_{ik}/RT)$。设 λ 是某序列中可选择字符数，$m_1 \ldots m_\lambda$ 是它们的单体前体的克分子浓度。模板字符 i 被其补体 j 识别的品质因子可以表达为：

$$q_{ij}=\frac{m_j K_{ij}}{\sum_{k=1}^{\lambda} m_k K_{ik}}$$

如果识别包括了协同作用，对各种最靠近的近邻的结合，这些因子可以不同。从不同长度的低聚核苷酸所获的实验数据(可参照几年前的实验室工作)指出，平均 q 值在0.99左右，这依赖于相对的 AU 及 GC 含量。这些值涉及到无酶识别的优化条件。改变物理环境不可能明显地提高它们，物理环境改变会同时影响互补性以及竞争相互作用。只有适合于该互补配对的独有的专--

识别，才能够导致深刻的改进。例如，与无酶识别相比，当$Q\beta$复制酶存在时，识别误差率减少了n个数量级。

错误复制的类型有三种，即字符的取代、缺失和嵌入，所有这些已在试管实验中得到证实。后面两种即缺失和嵌入，也可以由某些人为的诱变因素造成。特别是如果在进化的后期阶段涉及到已编码信息的翻译时，通常它们的结果更为严重。它们导致移码，从而可以毁灭所有后继的信息。对于所有那些效应，品质因子Q能够从形式上给予解释。

关于这种互补复制的线性模型，我们有如下结论：给定现实的字符品质因子(q_{ii})的值，我们将会知道，在某条已选定的聚合链中能够重复积累的信息容量太低了，以至难以允许某个酶机构的任何编码，这种编码是翻译和进一步增进复制所需要的。只有通过不同链的协同复制，才能取得充分大的容量。协同复制需要特定的耦合因子，这些因子具体说来很可能是蛋白质结构。

5.3 循环催化

蛋白质的三维结构具有巨大的功能容量，证明下面这个问题是有道理的：这些化合物是否不能把它们自己组织成为自复制的功能实体？

蛋白质结构肯定不具备如下的内在性质：它能够保护诸如自我指令和互补指令之类的事物。另一方面，业已发现某些酶复合体（公认为有高度复杂的结构）中的氨基酸是以十分确定的序列连结起来的，此种序列最长可在一条单链中包含15个之多的氨基酸残基。另一些酶可以使那些低聚核苷酸连结在一起，成为功能蛋白质结构。这种相当专一的、独特的蛋白质系综，如

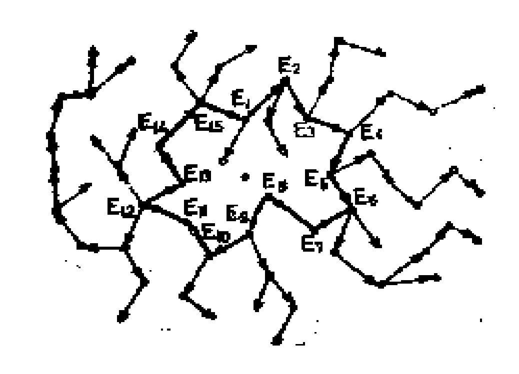


图 3-6　包括一个封闭环的蛋白质催化网络

在这个环上的任何酶 E_i，都是由前体借助酶 E_{i-1} 产生的。

果它形成如图3-6所示的催化循环，便可能成为自我复制的。

如果反应物(即氨基酸和低聚肽)大大超过了作为催化循环代表的合成酶，描述循环网络的速率方程便可线性化。类似于互补指令的模型，各个循环的竞争动力学性质可以用如下形式的矩阵表示：

$$\begin{pmatrix} -D_1 & 0 & \cdot & \cdot & \cdot & 0 & +F_1 \\ +F_2 & -D_2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & +F_3 & -D_3 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & +F_n & -D_n \end{pmatrix}$$

(其中 F_i 相应于第二个模型中的 A_iQ_i)。

这个矩阵具有 $k-1$ 个负的本征值，如果 $\Pi_k F_k > \Pi_k R_k$，可能还有一个正的本征值。而且，这个正的本征值包含了催化耦合常数的等比中项，它标征了一个单循环的选择价值，这个单循环可以与其他循环系综竞争。

初看起来，这样的系统对于我们的问题是相当有趣的，因

为它似乎能够集中起惊人数量的信息。然而，它存在的两个缺点会严重地限制进化优化作用的任何进步。

(1)该循环决不能摆脱寄生分支，这些寄生分支积累并终将使得所有竞争性结构连结起来(参见图3-7)。

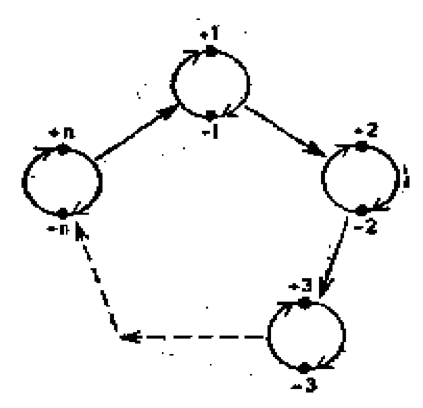


图 3-7　复制的RNA链的"超循环"

各个链系综的互补复制循环，都由于同前一个系综的一条链(或两条链)的特定相互作用而得到强化。这种在不同系综之间的联接，是循环闭合的。

(2)为了被复制，任何误差都必须传遍整个循环。换言之，它必须产生一定的误差序列，所有误差都具有导致它们自复制的独特性质。向更高催化效率适应的每一步骤，则会包括一个全新的循环起源，它是革命而不是进化。在这种系统中，进步将是极慢的，因为所需的误差巧合必须被看作非常稀少的事件。进化需要指令性质，这些性质是系统所固有的而不是某些复杂相互作用的独特结果。

5.4 超循环

迄今所考虑的模型，都是只在复杂性程度上有差别的线性反应系统。它要求：

(1)固有的自我指令或互补指令，以及

(2)实际上容量不受限制的可适应的催化能力，只有在几种反应作用物(很可能属于不同类的大分子，即核酸和蛋白质)的协同作用下，才能实现。协同行为由本质上非线性的反应机制反映出来。

图3-8表示一个非线性模型的简单例子，它只涉及到一类大分子——核酸。长度有限的互复制互补链的初级循环系综，可以通过某种循环催化耦合而特别地互相增强。这种催化过程的

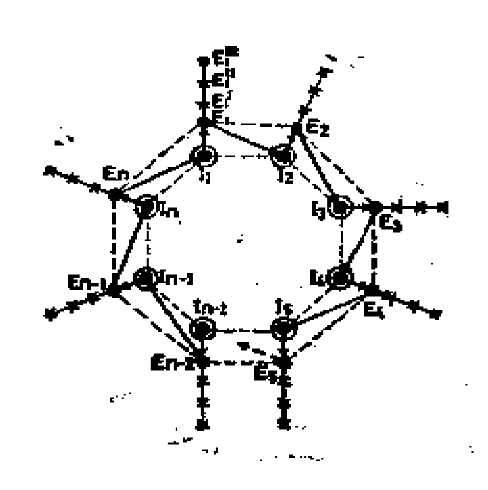


图 3-8　包括核酸和蛋白质的"催化超循环"

I_i 代表了 RNA 链的一个互补链系综。E_i——由 I_i 的"正"链编码——是一个蛋白质，它对后继的信息载体 I_{i+1} 提供功能联接。这些联接必须封闭起来，即最后有一个蛋白质 E_n，它促进 I_1 的形成。促进可以包括特定的复制、阻遏或抵制分解。I_i 的信息只有一部分可能被提供耦合 $E_i \to I_{i+1}$ 的因子所利用，其余部分可以包括一般性的效用(聚合，翻译，代谢功能，等等)。不过，读出 I_i 中的全部信息，必须由 E_{i-1} 中的一个共同的耦合因子来起动。

超循环由两组非线性方程来描述，其中一组用于信息载体 I_i，另一组用于功能 E_i 的载体。I_i 的产生项包含 I_i 和 E_{i-1} 两者的群体变量的表达式，而 E_i 的产生项由 I_i 的群体变量来决定。这些表达式以模板指令酶促聚合过程的随机理论结果为基础。对各种循环的计算借助于舒斯特建立起来的计算机程序。

循环等级组织可以称作“超循环”。图3-8所示的特殊系统是假设的。尚未找到明确的实验证据，以表明核酸能够施行这样的催化影响，即能特别地增强复制的速度和精确度或抵制分解。

任何系统的进化意义，取决于它所能产生的进入下一个较高层次的能力。对于图 3-8 中所示的系统来说，下一个较高层次将是利用蛋白质的更大的功能容量。这就要求核苷酸密码子对氨基酸有某种固有的分配关系，即密码和翻译的起源。而且，迄今尚未找到实验证据，它能证明核酸(不借助于蛋白质)能够促进这种分配。这并不排除存在某种尚未被认识的高级机制，但是它肯定极不利于任何直接的氨基酸-密码子关系的存在。

假定在密码子和反密码子或连接物和氨基酸之间，不存在任何内在的关系。则是否存在这样一个有限的概率，即以先前不受约束的蛋白质结构为媒介，使转译可能起源于随机分配的涨落中？这是一个合理的问题，因为我们有证据表明，蛋白质原则上能够执行这种功能，在一个随机集合中任何可能的分配以有限的机会出现。

假定我们打算翻译 λ 个字符：

$$a, b, c, \ldots,$$

借助于相应数目的连接物：

$$A, B, C, \ldots,$$

于是有λ^2种可能的分配，即任一字符对任一连接物。每一个分配可由一特定催化剂带来，在某个蛋白质随机集合中可以找到每一种这样的催化剂，它们具有相同的先验概率。我们必须选择的是 λ 种催化剂的一个集合，这些催化剂促进某种独特的分配，每一类型字符与且只与一种连接物相联系，而且反之亦然。在总计$C_\lambda=\binom{\lambda^2+\lambda-1}{\lambda}$的集合中，出现 $\lambda!$ 个这种独特的集合。

将所有 Q^2 个可能的分配(即"有重复的结合")组合成 λ 个元素的组从而获得 C_λ。

现在我们设想一些分隔间，它们足以容纳 λ 个这种催化剂，这类催化剂促进某几种分配。同时，分隔间应该包含等当数量的随机字符序列，这些字符依据任何一种翻译产生出功能结构，我们又能从这些结构中找到用于分配的λ种催化剂。我们怎样才能够常常预期某个空间隔区，在其中：

(*a*) 一个独特的催化剂集合引起明确的翻译，以及

(*b*) 这种翻译的产物增强了相同类型的分配。它将是 $\frac{C_\lambda^2}{\lambda!}$ $=\frac{[(\lambda^2+\lambda-1)!]}{[\lambda!]^3[(\lambda^2-1)!]^2}$ 个空间隔区中的一个隔区(参见表 3-3)。

表 3-3

λ	2	4	8	20
$\frac{C_\lambda^2}{\lambda!}$	~50	~7×10^5	~3×10^{15}	~10^{50}

注意，只要求功能的某种相符合而不是序列的相符合，否则这些数目将是惊人的巨大。就是这种来自暂态事件的符合产生出基因型信息的语义学。这种涨落的出现究竟有多频繁，取决于假设的空间隔区的"尺寸"，即取决于在任一随机蛋白质结构的集合中找到分配功能的机会。任何偏离一致先验概率的倾向，都将增加这种机会。

这些数字(表3-3中已给出一些例子)表明，对任何在8或10以上的λ值，随机开端变得极不可能。当然，作为λ=20时出现的10^{50}个选择，在任何现实的时间间隔内绝不可能容纳在地球上。某个前体密码兴许从氨基酸分类开始，而不是从一个个分

配开始。如果就原始蛋白质的结构而言，四类差别，诸如非极性，极性(但处于电中性)、正电性和负电性，可能对应于某种起码要求。功能上的需要可能会增加种类数，以包括催化活性基团，例如具有特殊的酸性或碱性的基团。增至八类时将相似于运用二元字符的三联体密码的最大容量。确实，遗传密码表强烈地提示预先存在某种类型的分类。

翻译的出现，带来了一个表现型表达的新的方面。然而，如果典型的基因型系综以这样一种方式组织它自己，即它作为一个整体被选择从而能保存它的信息内容，那么只有这样它才能在继续进化发展中被利用。图3-9模式所表示的模型，其中包括核酸和蛋白质的协同作用，并能使生物合成翻译器的选择性稳定化。它是以如下性质来区分的：

(1)每一个循环作为一个整体，具备自促进的生长性质。由于非线性放大项，群体变量中的奇点可能会在有限的时间内出

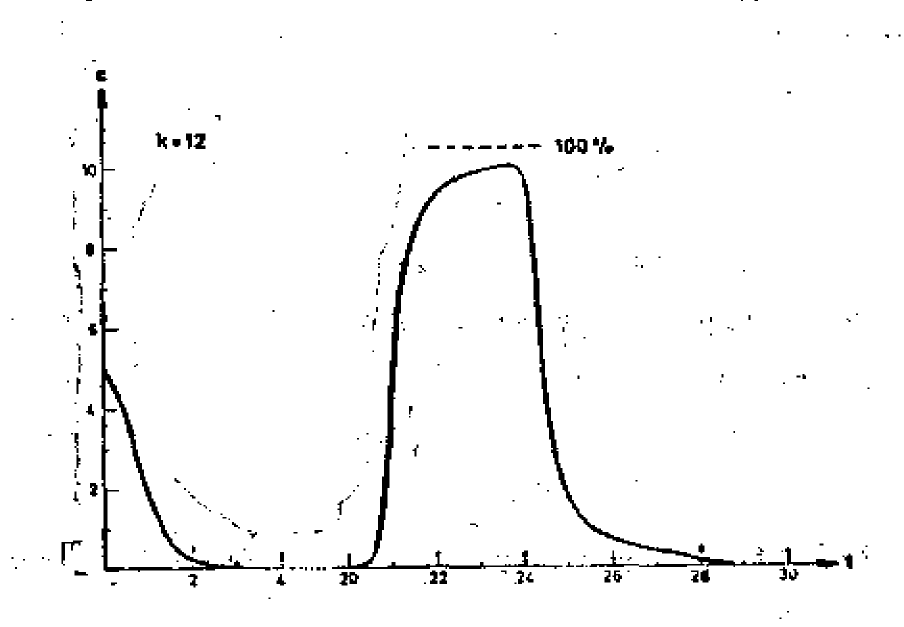


图 3-9　一个(浓度对时间)解的例子　这个解描述了一个12元超循环(即包含 $I_1, \cdots, I_{12}; E_1, \cdots, E_{12}$)中浓度的循环变化。假定所有步骤的速率常数相等。时间单位以速率常数的倒数给出。

现。（在图3-8的说明中，描述了I系统和E系统的非线性速率方程的类型。）

（2）在生长控制下（例如，恒定流量或总反应力），各个独立的循环将为选择而竞争。非线性选择的结果极为明显的，类似于“全或无”行为。

（3）如果出现一类大分子成分（I或E）大大过量，那么可以假定该系统的准线性行为。在非线性范围内，稳态解包含了无阻尼振荡。只要超循环中每类的组分不多于3个，简正型分析（如同舒斯特按钱德拉斯卡程序实现的）便揭示出稳定的定常态。组分为4个时，发生无阻尼谐振荡（纯虚根），而当每一类的组分超过4个时，则出现不稳定解。任何对稳态的偏差都将放大，但是由于生长限制，最后又倒转过来。不同组分的浓度变化类似于环绕循环的波形，图3-9显示了这种例子。

（4）消去时间变量和数值积分，导致在一个k维空间（因为每类循环中有k种组分）的相图。在2维子空间的投影产生出积分曲线，这些曲线趋近于$k>4$某个极限循环。这个ν维子空间由相继的两个组分I_i和$I_\lambda+1$的群体变量来估量，对于大的k值，该极限循环退化成为三角形，类似于几乎是矩形的浓度波。对于选择的速度和明确性，这种行为是重要的。

（5）此种特殊的非线性性质，使这种系统（a）在随机的有效程度上能够利用非常小的选择优势，（b）非常迅速地进化，以及（c）在给定的超循环已放大到某种群体水平之后，不再容忍任何新竞争者。因此，即使随机开始任意的密码分配，密码的普适性以及所有聚合体结构的一致的手性，都将由于新的超循环成核的阈值很高而得到保护。这种非线性竞争，将更加强烈地不利于任何始于低群体水平的“新生者”。

（6）循环耦合提供了充分大的信息容量，这是单链决不可能

达到的(参照N_{max}),即令此时存在(仍是轻微适应了的)酶。计算表明，此循环将自动地节省它的规模，直至达到某些功能需求的地步，这些需求保证它的存在。发生在一条单链I_i中的基因型变化可以立刻被利用，如果它们加强了与E_{i-1}的相互作用的话。它们不必围绕该循环传播进一步的变化，而这是线性系统中的直接耦合所必要的。

(7)超循环闭环之外的寄生分支(即与其他I—循环的耦合)不为该系统所容忍。那些分支要么是干扰超循环的起源，要么系统不对它们进行选择。如果它们在该循环已达到一定的群体水平之后出现，它们就完全没有机会作为非线性选择行为的某种结果生长起来。

(8)要使分支进化,只有一种方式。它们必须是环内I链的一部分。然而，它们或许会有比编码耦合因子E_i所需要的更多的信息，这个耦合因子连接I_i与I_{i+1}。无论何时单个I单元联结在一起，这将自动地导致基因组的操纵子结构。今天所遇到的细胞，其染色体展现了这种操纵子结构。它们由基因单元组成，通常包含几种基因：在诱导物及抑制基因(二蛋白质因子)的控制之下，这些基因协同地起作用,诱导物及抑制基因与DNA的控制部位相互作用。这类分支的进化,需要把$(1-\bar{q})$调节到小于10^{-3}。由聚合化、翻译和代谢酶所表达的一般效用的全部功能，终将在这些分支中出现,而E_i是特定的耦合因子——诸如通过保护、抑制—去抑制,或者任何其他类型诱导起作用的"外壳蛋白"。

(9)出现在操纵子分支中的选择优势的共化利用,终归需要几种分隔间。关闭进分隔间中将保证群体抵御出现在单物种中的有害突变。由于稀释问题，它可能还在非常早的阶段就相当重要了。密码和翻译器的集结，也许实际上已经出现在分隔化的系统中。这些早期的分隔间应该是相同开放的结构，例如象

核糖体或奥巴林提出的团聚体那样的复合体。在模拟前生物条件的实验室实验中，反复发现了一些微球体，这些微球体可以在较后的阶段中，在保证某种已有的系综方面提供进一步的优势。

(10)一个分隔化的系统，由把所有单个I单元联结在一起成为一条稳定链(可能是某种双链DNA)，终将也不得不"个体化"。把核酸链的"粘性"的两端联结起来的功能，也许曾在其中一条分支上进化。目前我们对那些连接酶已有了很好的了解。为细菌染色体确立的环状基因图(参见图3-10)，可能曾经以这种方式进化的。另一个结果是不同操纵子的随机排列。(由基因I_i引起的)耦合因子可以与其目标基因(I_{i+1})相距颇远，而只有对于封闭在给定操纵子中的协调功能才可能接近相遇。上面提到的后两种效应，不是从特殊的非线性微分方程系统的数学结构

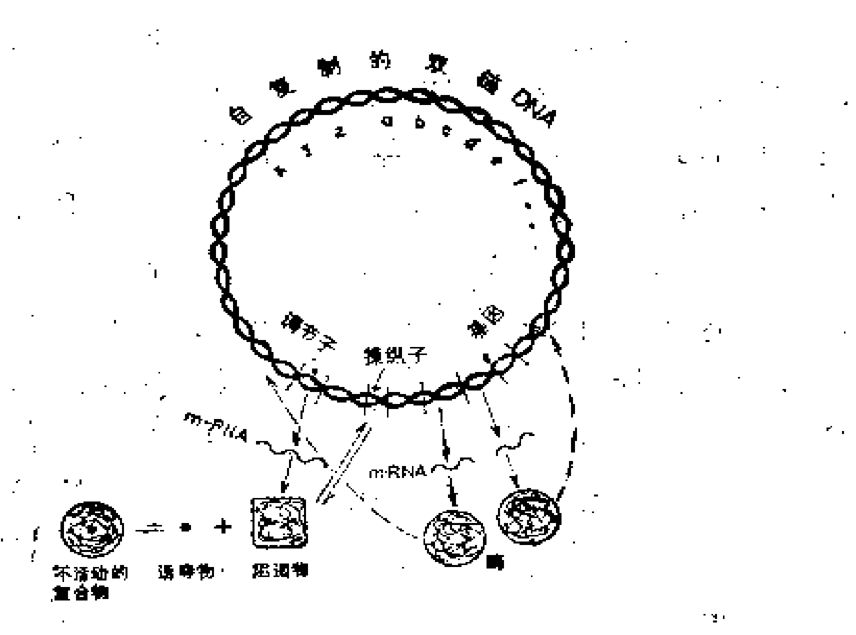
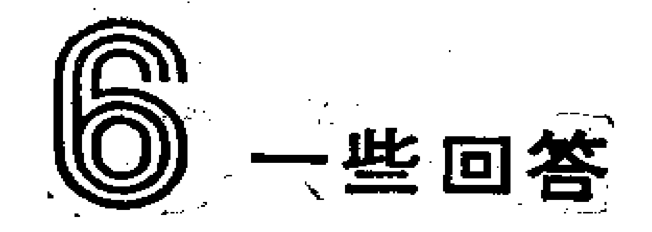


图 3-10 超循环的最终进化产物，类似于细菌细胞的环状染色体的基因和操纵子的结构。单个信息载体I_i由连接酶(它已闭合了多肽链的开放端)连接在一个环上，从I_i上可导出基因。在这种环状基因组中，基因I_i和I_{i+1}(通过E_i因子耦合)不必相邻。

直接推导出来的结果。不过，任何此种提供了隔离化和个体化的可能性的系统，都将利用它，因为它具有生存优势。它们表明，超循环不是末日，而是要导致新的进化水平，这或许可以被认为与细菌细胞的原始前体一致。



6 一些回答

本文是从如下一些问题开始的：

我们能够定义生命标准吗？

它们是以已知的物理学规律为基础的吗？

“生命”现象是注定要出现的吗？

我们了解了(大)分子的前体以及它们出现的顺序吗？

虽然上述这些问题中的每一个问题是以这样一种形式提出的，因而都可简单地以“是”或“否”来回答，但在结论中，我的回答将稍微详细一些。

让我们从三条“经典的”生命标准入手。

(1)自我复制——以保存生物信息，尽管有平稳的破坏。

(2)变异性和选择——以扩大信息量，为某种价值标准造成某种倾向。

(3)代谢作用＝自由能流——以补偿平稳的熵产生，这与同时满足标准(1)和(2)相联系，而且由此阻止了总观平衡。

利用某些特殊的分子性质，并保持一定的物理条件，就可以满足这些标准。它们是生命的必要前提。它们导致物质系统的自组织。不稳定性平稳地减少了状态可能的可选择数，有时则达到唯一确定的最优条件。这种在一定集合中的可选择数的

减少，相应于某种"信息增加"①。尽管拷贝选择本质上是随机事件，最优化过程在确定的涨落限度内仍然是决定论的。可以提出一个价值原理，它支配着选择最优化过程。它以一般的稳定性标准为基础，与非平衡热力学的原理(详见普里戈金的讲义)一致。

这些标准肯定是必要的；但是，对于描述"生命"现象它们是否是充分的？如果我们指的是生命这个术语，我们则必须把它理解为我们今天实际上遇到的现象，例如，细菌、植物或动物等形式。正如前面的讨论所表明的，这些标准甚至不足以适当地描述各种模型。为保证从任何给定水平向下一个较高级水平(由新的特征所区别)的无止境的进化，必须满足另外的更专门的要求。当然，模型只能揭示在给定水平上的主要的潜在可能性并揭示其限度。应该明确强调的是，上述的模型并不打算要模拟进化的历史过程(在任何如同生命细胞那样复杂的东西产生之前，进化实际上也许已经通过了许多中间过程)。不过，它们告诉我们，哪些可能的选择是被排除掉的。而且，它们说明了一个系统要成为下一层次的真正前体必须带有哪些性质。另一方面，进化不必是一个完全单调的过程，而可以包括一些不连续性即总观不稳定性，于是我们对此仍然可使用"起源"这个术语。作为超循环基础的知识，基本上"收集"来自蛋白质和核酸的先前自由结构，因此这类知识正是起源于这个层次。

对于理解生命现象，已知的物理学规律提供了全部的基础吗？

目前我们能够做到的是，援引反例来反驳对这类主张的否定看法。换言之，对于认为已知物理学规律不足以说明标志生

---

① 按照伦伊的定义。

物体现象的任何主张，我们都能够运用物理模型去反驳它们。超循环就是一个例子。

这种说法至少对非常原始形式的生命是正确的，我们的讨论(即关于"起源"的讨论)就限制在比较原始的生命形式的范围。没有理由认为这种说法对随后的任何进化阶段不应该成立，

(*a*)有性复制——其特征是在不同的物种之间的信息交换，从而可以更快地利用优势。正如雅各比强调的，在这种水平上，死亡成了进化的前提。

(*b*)分化和形态发生——是在细胞层次上的自组织过程，但在分子机制的控制下，这些分子机制以本文讨论过的特征著称。

(*c*)通过神经系统与环境的通讯，再一次涉及几个阶段。范围从(1)简单控制，到(2)某种新的"经验性"信息的自组织，并受"二级"评估系统的指导，最后到(3)通过评估响应，抽象地再现出信息和选择性的集中，我们将它与意识这个术语联系起来了。

这些20世纪物理学的新的基本思想，并不是产生于知识的缺乏，而是来自于实验事实和逻辑结论之间的不一致，这些逻辑结论是以先前公认的概念为基础的。对于我们称为"生命"的这种物质组织的高级层次以及其它新奇的无形结果，我们的知识仍然十分缺乏。然而，在这个阶段，它只不过是"知识缺乏"，而不是与现在的物理学概念"不一致"。